细胞的分子生物学

原书第四版

MOLECULAR BIOLOGY OF THE CELL

（上册）

BRUCE ALBERTS 等著

张新跃　钱万强　等译

科学出版社
www.sciencep.com

(Q-1726.0101)

细胞的分子生物学

原书第四版

MOLECULAR BIOLOGY OF THE CELL

高教出版中心 农林与生命科学分社
联系电话：010-64030233
http://www.sciencep.com
e-mail：bio@mail.sciencep.com

ISBN 978-7-03-017214-3
9 787030 172143 >

销售分类建议：生命科学/细胞生物学

全套定价：168.00 元(含光盘)上 下册

国外生命科学优秀教材译丛

Molecular Biology of The Cell
细胞的分子生物学

(原书第四版　上册)

〔美〕B. Alberts 等　著

张新跃　钱万强 等　译

科学出版社

北　京

图字：01-2003-7128

Molecular biology of the cell/Bruce Alberts... [et al.]. —4th ed.

ISBN 0-8153-3218-1

© 2002 by Bruce Alberts et al.

All rights reserved. No part of the publication may be reproduced, stored in a retrieval system or transmitted in any form or by any means—electronic, mechanical, photocopying, recording, or otherwise—without the prior written permission of the copyright holder.

Authorized translation from english language edition published by Garland Science, an imprint of Taylor & Francis Books, Inc.

本书封面贴有 Taylor & Francis 集团防伪标签，未贴防伪标签属未获授权的非法行为。

图书在版编目(CIP)数据

细胞的分子生物学/（上）（美）艾伯茨（Alberts, B.）等著；张新跃等译. —北京：科学出版社，2008

（国外生命科学优秀教材译丛）

ISBN 978-7-03-017214-3

Ⅰ. 细… Ⅱ. ①艾…②张… Ⅲ. ①细胞生物学-教材②分子生物学-教材 Ⅳ. ①Q2②Q7

中国版本图书馆 CIP 数据核字(2006)第 045784 号

责任编辑：周 辉 彭克里 席 慧/责任校对：张 琪

责任印制：张克忠/封面设计：耕者设计工作室

科学出版社 出版

北京东黄城根北街 16 号

邮政编码：100717

http://www.sciencep.com

双青印刷厂 印刷

科学出版社发行 各地新华书店经销

*

2008 年 8 月第 一 版 开本：787×1092 1/16

2008 年 8 月第一次印刷 印张：103 3/4

印数：1—4 000 字数：2 430 000

定价：168.00 元（含上下册，光盘）

（如有印装质量问题，我社负责调换（双青））

译 者 名 单

主译： 张新跃　钱万强　舒　畅　张远涛

审校： 张新跃

翻译人员名单（按汉语拼音顺序排名）：

董　娜　韩　冬　何善平　鲁　洁　卢　路　钱万强

舒　畅　武　晔　杨勇飞　张　辰　张丹丹　张晓明

张新跃　张　旭　张远涛

译 者 的 话

在2005年接到科学出版社委托我们翻译生命科学领域的权威教材——细胞的分子生物学（*Molecular Biology of The Cell*）一书的任务时，确实有些犹豫：因为工作量浩大，难度不小，要对图书翻译质量负责，时间又很紧张。经过一番激烈的思想斗争之后，再加上盛情难却，我们最终接收了这一艰巨的任务。随后，我们组织了一批年轻的生命科学才俊，并根据各自的擅长领域和兴趣，分配了翻译任务，由此开始了漫长的翻译和校对过程。

细胞的分子生物学是生命科学领域非常重要的一门基础学科。Bruce Alberts等主编的*Molecular Biology of The Cell*一书是全世界公认的一部非常优秀的教材。科学出版社引进这一教材并将其翻译成中文版，确实是明智之举，但愿该书的出版有助于推动我国生命科学的教学和科研。因为译著只是译者对原著的理解，这与译者的翻译和学识水平有关，因此译文很难完全忠实于原文，有时甚至还有错译和误译之处，尽管我们在翻译和校对过程中竭力避免出现此类现象。

参与本书翻译工作的译者分工如下：张新跃（前言、致谢、致读者、词汇表、第11章、第23章至第25章）、钱万强（第1章）、何善平（第2章和第15章）、卢路（第3章）、张旭（第4章至第7章）、张丹丹和鲁洁（第8章）、舒畅（第9章和第10章）、武晔（第12章和第22章）、董娜（第13章和第14章）、张辰（第16章）、张晓明（第17章）、韩冬（第18章）、张远涛（第19章和第20章）、杨勇飞（第21章）。

由于译者的翻译水平和学识水平所限，错译和误译在所难免，敬请读者在阅读过程中不吝赐教。

译者

2008年3月31日于美国马里兰

使 用 说 明

因本书为黑白印制，考虑到方便读者使用，特将全书所有图片放入所附光盘中。书中所有译文中关于颜色的描述可参照光盘使用。

目　　录

第Ⅰ部分　细胞导论

1　细胞和基因组 …… 3

地球上细胞的共同特征 …… 3

所有的细胞都以同样的线性化学密码（DNA）形式储存遗传信息 …… 3

细胞通过依照模板的聚合作用复制遗传信息 …… 4

所有的细胞都将其部分遗传信息转录成共同的中间体（RNA） …… 5

所有细胞都将蛋白质用作催化剂 …… 7

所有细胞都以相同的方式将 RNA 翻译成蛋白质 …… 8

相应于一种蛋白质的遗传信息片段就是一个基因 …… 9

生命需要自由能 …… 10

所有的细胞都是有着相同分子建造材料的生化工厂 …… 11

细胞外被覆一层细胞膜，营养物质和废弃物必须通过细胞膜进出细胞 …… 12

细胞的生存仅仅需要不到 500 个基因 …… 12

小结 …… 14

基因组的多样性及生命树 …… 14

细胞可由多种自由能源提供能量 …… 14

一些细胞可以为其他细胞固定氮及二氧化碳 …… 16

原核细胞间的生化多样性是最为广泛的 …… 16

生命树有三个基本分支：细菌、古细菌和真核生物 …… 18

一些基因演化迅速，另一些则十分保守 …… 19

大多数细菌和真细菌含有 1000～4000 个基因 …… 20

新基因产生于先前存在的基因 …… 21

基因复制引起单个细胞内相关基因家族的出现 …… 21

基因可以在两个物种之间相互转移，这种现象在实验室和自然界都可以发生 …… 23

物种中基因信息的水平交换是由性引起的 …… 25

基因的功能常能根据其序列而推测 …… 25

生命树上三个基本分支间有 200 多个基因家族相同 …… 26

突变揭示基因的功能 …… 27

分子生物学家将焦点对准了大肠杆菌 …… 27

小结 …… 28

真核生物的遗传信息 …… 29

真核生物起源于捕食生物 …… 29

真核细胞起源于共生体 …… 31

真核生物有着杂和的基因组 …… 33

真核生物的基因组非常庞大 …… 33

真核基因组中有着丰富的调控 DNA …… 34

基因组决定多细胞发育的进程 …… 35

许多真核细胞以单细胞的形式存在，即原生生物 …… 36

酵母是最小的真核模式生物 …… 36

机体中所有基因的表达水平都同时受到监控 …… 38
拟南芥——300 000 多种植物的模式物种 …… 39
动物细胞的代表物种：线虫、果蝇、小鼠、人 …… 40
果蝇的研究为脊椎动物发育学提供了钥匙 …… 40
脊椎动物基因组是重复复制的产物 …… 42
基因冗余对于遗传学家来说是个大问题，但却为进化中的物种提供了机会 …… 43
小鼠是哺乳动物的模式生物 …… 44
人类可以报道自身的特性 …… 45
精确来说我们所有人都是不同的 …… 45
小结 …… 46
2 细胞的化学和生物合成 …… 48
细胞的化学组成 …… 48
细胞是由几种原子构成的 …… 48
最外层电子决定原子的作用方式 …… 51
电子的获得和丢失形成离子键 …… 52
共价键是通过共用电子对形成的 …… 53
存在几种不同类型的共价键 …… 54
专题 2-1 生物分子中常出现的化学键和化学基团 …… 56
原子的行为常常表明它的半径似乎是固定的 …… 58
专题 2-2 水及其对生物分子行为的影响 …… 59
水是细胞中含量最丰富的物质 …… 61
某些极性分子在水中形成酸和碱 …… 61
4 种非共价相互作用帮助细胞内分子结合 …… 62
细胞是由碳化合物构成 …… 64
细胞含有 4 大类主要家族的有机小分子 …… 64
专题 2-3 结合大分子的主要类型的弱非共价键 …… 65
糖类物质为细胞提供能量来源，也是多糖的亚单位 …… 67
专题 2-4 细胞内常见糖的一些类型的概述 …… 69
脂肪酸是细胞膜的组成物质 …… 71
氨基酸是蛋白质的亚单位 …… 72
专题 2-5 脂肪酸和其他的脂类 …… 73
核苷酸是 DNA 和 RNA 的亚单位 …… 77
专题 2-6 核苷酸的概括 …… 79
拥有显著特征的大分子在细胞化学中占据主要地位 …… 81
非共价键不仅决定了大分子的精细形状，而且决定了它与其他分子的结合 …… 82
小结 …… 83
催化作用和细胞利用能量 …… 84
细胞代谢是由酶组织的 …… 84
细胞释放的热能使得生物有序性成为可能 …… 86
专题 2-7 自由能和生物反应 …… 88
光合生物利用阳光合成有机分子 …… 91
细胞通过氧化有机分子获取能量 …… 91
氧化与还原涉及电子转移 …… 93
酶降低了阻遏化学反应的障碍 …… 94
酶是怎样找到底物的——迅速扩散极其重要 …… 95

自由能变化决定反应能否发生 …… 98
反应物浓度影响 ΔG …… 98
对于连续反应，ΔG° 值是可加和的 …… 100
活化的载体分子对于生物合成必不可少 …… 102
活化载体的生成与能量上有利的反应偶联 …… 102
ATP 是最广泛适用的活化载体分子 …… 103
储存于 ATP 中的能量通常用于两个分子的接合 …… 103
NADH 和 NADPH 是重要的电子载体 …… 105
细胞内还有许多其他的活化载体分子 …… 107
生物聚合物的合成需要输入能量 …… 108
小结 …… 111
细胞怎样从食物中获取能量 …… 112
食物分子分三个阶段分解产生 ATP …… 113
糖酵解是生成 ATP 的中心途径 …… 115
发酵使得在无氧条件下能够生成 ATP …… 115
专题 2-8　糖酵解途径中 10 个步骤的详细内容 …… 117
糖酵解过程证实了酶是如何将氧化放能与能量储存偶联起来的 …… 119
糖和脂肪都在线粒体分解为乙酰 CoA …… 122
柠檬酸循环使乙酰 CoA 氧化成 CO_2，生成 NADH …… 123
电子转移推动细胞内大多数 ATP 的合成 …… 125
专题 2-9　完整的三羧酸循环 …… 127
有机体使用特殊的仓库储存食物分子 …… 129
氨基酸和核酸参与了氮循环 …… 132
许多生物合成途径起始于糖酵解作用或柠檬酸循环 …… 132
代谢受到组织和调节 …… 133
小结 …… 135
3　蛋白质 …… 138
蛋白质的形状和结构 …… 138
蛋白质的形状特异性决定于其氨基酸序列 …… 138
专题 3-1　蛋白质中的 20 种氨基酸 …… 141
蛋白质折叠为能量最低的构象 …… 144
α 螺旋和 β 折叠是常见的折叠模式 …… 145
专题 3-2　显示了 4 种不同的描述 SH2 结构域（真核细胞中有重要功能）的方式 …… 147
结构域是蛋白结构的一个基本单位 …… 149
可能的多肽链中只有少数是有用的 …… 149
蛋白质可以形成有限的折叠模式 …… 151
同源序列搜索可以鉴定亲缘关系 …… 152
计算机技术可以将氨基酸序列归类为已知的蛋白质折叠模式 …… 153
一些被称为模块的蛋白质结构域，形成了很多不同蛋白质的一部分 …… 154
人类基因组编码一套复杂的蛋白质，很多仍然是未知 …… 155
大的蛋白质通常包含不止一条多肽链 …… 156
一些蛋白质形成长的螺旋状纤维 …… 157
一个蛋白质分子可以形成长的、纤维状结构 …… 159
共价交联稳定胞外蛋白 …… 161
蛋白质分子通常作为大的结构分子的亚基 …… 161

细胞内的很多结构是自组装的 …… 163
复杂生物结构的形成需要辅助因子的帮助 …… 165
小结 …… 166
蛋白质的功能 …… 167
所有的蛋白质都可以结合其他分子 …… 167
蛋白质构象的细节决定了其化学性质 …… 168
蛋白质家族成员序列比对可以发现重要的配体结合位点 …… 169
蛋白质通过多种类型的接触面相互结合 …… 170
抗体的结合位点是高度可变的 …… 170
结合能力由平衡常数来衡量 …… 171
酶是高效性和高度专一性的催化剂 …… 172
底物的结合是酶促反应的第一步 …… 174
专题 3-3　用来研究酶的一些方法 …… 175
酶通过选择性的稳定转换状态加速反应 …… 177
酶可以同时产生酸催化和碱催化作用 …… 177
溶菌酶揭示了酶是怎样发挥作用的 …… 177
与小分子的高亲和性赋予了蛋白质额外的功能 …… 181
多酶复合物帮助增加细胞代谢速率 …… 182
酶的催化活性可以被调节 …… 183
别构酶有两个或多个结合位点相互作用 …… 184
两个配体如果其结合位点是偶联的，将会相互影响各自的结合性 …… 184
对称的蛋白质分子的组装产生协同别构转换 …… 186
原子水平上了解天冬氨酸转氨甲酰酶的别构转换 …… 187
磷酸化会导致蛋白质的很多变化 …… 188
真核细胞中有一大类蛋白激酶和蛋白磷酸（酯）酶 …… 188
Cdk 和 Src 蛋白激酶的调控显示了一个蛋白质是如何作为一个微芯片起作用的 …… 191
蛋白质结合和水解 GTP 是普遍存在的细胞调节因子 …… 192
调控蛋白通过决定 GTP 或者 GDP 结合来控制 GTP 结合蛋白的活性 …… 193
大蛋白质的运动来自于小蛋白质 …… 193
动力蛋白负责细胞中的大运动 …… 196
膜结合转运蛋白利用能量将分子输送过膜 …… 198
蛋白质通常形成巨大的复合体，以蛋白质机器的形式发挥功能 …… 199
细胞功能的基础是复杂的蛋白质相互作用网络 …… 199
小结 …… 201

第 II 部分　基本遗传机制

4　DNA 与染色体 …… 205
DNA 的结构和功能 …… 207
DNA 分子由两条互补的核苷酸链组成 …… 207
DNA 的结构提供了一种遗传机制 …… 210
真核生物中，DNA 围在细胞核内 …… 212
小结 …… 213
染色体 DNA 及其在染色质纤维中的包装 …… 213
真核生物 DNA 包装成一套染色体 …… 213

染色体含有长串的基因 …… 214
人类基因组的核苷酸序列揭示了基因在人体内是如何排列的 …… 217
相关生物 DNA 之间的比较揭示 DNA 序列中存在保守区域和非保守的区域 …… 220
染色体存在于细胞生命过程中的不同阶段 …… 221
每个形成线性染色体的 DNA 分子都必须含有一个着丝粒、两个端粒和复制起点 …… 222
染色体中的 DNA 分子高度凝聚 …… 224
核小体是真核生物染色体结构的基本单位 …… 224
核小体核心颗粒的结构揭示了 DNA 是如何包装的 …… 226
核小体在 DNA 上的位置由 DNA 柔性和其他 DNA 结合蛋白质决定 …… 227
核小体通常一起包装成一条致密的染色质纤维 …… 229
ATP 驱动的染色质重建装置改变了核小体结构 …… 231
组蛋白尾的共价修饰可以对染色质产生深远的影响 …… 232
小结 …… 235
染色体的总体结构 …… 235
灯刷染色体含有解凝聚的染色质环 …… 235
果蝇多线染色体排列成交替的带和间带 …… 237
在多线染色体中，带和间带都含有基因 …… 239
单个多线染色体的带能作为一个单位进行解折叠和重折叠 …… 240
异源染色质是高度组织的，通常会抑制基因表达 …… 242
染色体末端存在一种特殊形式的异染色质 …… 243
着丝粒也包装成异染色质 …… 246
异染色质可能提供了一种抵制移动 DNA 元件的机制 …… 249
有丝分裂染色体由处于最凝聚状态的染色质形成 …… 250
每条有丝分裂染色体都含有巨大结构域这种特征模式 …… 251
单个染色体占据间期细胞核内连续区域 …… 253
小结 …… 255
5 DNA 复制、修复以及重组 …… 257
DNA 序列的维持 …… 257
突变率极低 …… 257
蛋白质中许多突变都是有害的，并被自然消除 …… 258
据我们所知，低突变率对于生命是必需的 …… 259
小结 …… 259
DNA 复制机制 …… 259
碱基配对原则为 DNA 复制和修复提供了基础 …… 259
DNA 复制叉是不对称的 …… 261
DNA 复制的高度忠实性需要多种校对机制 …… 263
只有以 5′→3′方向进行的 DNA 复制能够对错误有效地加以校正 …… 265
一种特殊的核苷酸聚合酶在后随链上合成短 RNA 引物分子 …… 265
特殊蛋白质帮助打开复制叉前面的 DNA 双螺旋 …… 268
移动的 DNA 聚合酶分子通过一个滑动的环保持与 DNA 的连接 …… 269
复制叉上的蛋白质协同作用，形成一个复制机器 …… 271
一种链指导的错配修复系统清除了逃过复制机器检查的复制错误 …… 274
DNA 拓扑异构酶防止复制期间 DNA 乱成团 …… 275
真核生物和细菌的 DNA 复制过程相似 …… 276

小结 …… 279
染色体上 DNA 复制的起始与完成 …… 280
DNA 复制起始于复制起点 …… 280
细菌染色体只有一个 DNA 复制起点 …… 280
真核生物的染色体含有多个复制起点 …… 281
在真核生物中 DNA 复制只在细胞周期的一段时间内进行 …… 284
同一染色体上的不同区域在 S 期的不同时间进行复制 …… 284
高度压缩的染色质复制较晚，而压缩程度较小的染色质中的基因倾向于较早复制 …… 285
在一种简单真核生物芽殖酵母中确定的 DNA 序列充当复制起点 …… 285
一个巨大的多亚基复合物结合在真核生物复制起点上 …… 287
哺乳动物中指定复制起始的 DNA 序列曾一度难以确定 …… 287
新的核小体在复制叉后面组装起来 …… 288
端粒酶复制染色体的末端 …… 289
端粒长度受到细胞和生物体的调控 …… 292
小结 …… 293
DNA 修复 …… 294
如果不存在 DNA 修复，自发的 DNA 损伤将迅速改变 DNA 序列 …… 295
DNA 双螺旋容易修复 …… 296
DNA 损伤可以通过多种途径加以清除 …… 297
DNA 碱基的化学促进损伤的发现 …… 298
双链断裂被有效地修复 …… 299
在对 DNA 损伤的应答中细胞能产生 DNA 修复酶 …… 301
DNA 损伤延迟了细胞周期的进程 …… 302
小结 …… 302
一般性重组 …… 302
一般性重组由两个同源 DNA 分子间碱基配对互作指导进行 …… 303
减数分裂重组由双链 DNA 断裂引起 …… 304
DNA 杂交为一般性重组中碱基配对步骤提供了一种简单模型 …… 305
RecA 蛋白及其同源蛋白质使 DNA 的一条单链能够与 DNA 双螺旋的同源区域配对 …… 305
真核生物中存在 RecA 蛋白的多种同源蛋白质，它们每一种都特化后具有特异的功能 …… 308
一般性重组常常涉及一个 Holliday 连接的形成 …… 308
一般性重组能引起基因转换 …… 309
一般性重组事件在有丝分裂和减数分裂细胞中倾向于产生不同的结果 …… 311
错配校对防止两条不能配对的 DNA 序列之间发生任意重组 …… 313
小结 …… 314
位点特异性重组 …… 314
可移动遗传元件能够通过转座或者保守机制进行移动 …… 315
转座的位点特异性重组能将可移动遗传元件插入到任何 DNA 序列中 …… 315
DNA 转座子提供 DNA 断裂和连接机制进行转移 …… 317
有些病毒利用转座的位点特异性重组将它们自身转移到宿主细胞染色体中 …… 319
逆转录病毒样的逆转座子与逆转录病毒类似，但是缺少蛋白质衣壳 …… 319
人类基因组大部分是由非反转录病毒的逆转座子组成 …… 321
在不同生物中存在不同的占优势地位的转座元件 …… 321
基因组序列揭示出转座元件发生过转移的大概时间 …… 322

保守的位点特异性重组能可逆地重排 DNA …… 323
保守的位点特异性重组能够用来打开或关闭基因 …… 326
小结 …… 327
6 细胞如何解读基因组 …… 329
从 DNA 到 RNA …… 331
部分 DNA 序列转录成 RNA …… 331
转录产生与 DNA 一条链互补的 RNA …… 332
细胞合成几种类型的 RNA …… 336
编码在 DNA 中的信号告诉 RNA 聚合酶从哪里开始，又在哪里结束 …… 336
转录起始和转录终止信号是不同的核苷酸序列 …… 339
真核生物中转录起始需要许多蛋白质 …… 341
RNA 聚合酶Ⅱ需要通用转录因子 …… 342
聚合酶Ⅱ也需要激活蛋白、介导蛋白以及染色质修饰蛋白 …… 342
转录延伸在 DNA 中产生超螺旋张力 …… 345
真核生物中转录延伸与 RNA 加工紧密偶联 …… 347
RNA 加帽是真核生物前体 mRNA 第一种修饰 …… 348
RNA 剪接从新转录的前体 mRNA 中除去内含子序列 …… 350
核苷酸序列为在哪里进行剪接提供了信号 …… 352
RNA 剪接由剪接体完成 …… 353
剪接体利用 ATP 水解提供能量形成各种复杂的 RNA-RNA 重排 …… 353
对前体 mRNA 的有序影响有助于解释如何选择正确的剪接位点 …… 355
另外一套 snRNP 剪接了动物和植物中一小部分内含子 …… 357
RNA 剪接表现出显著的可塑性 …… 358
由剪接体催化进行的 RNA 剪接可能从自我剪接机制演变而来 …… 359
RNA 加工酶产生真核 mRNA 3′端 …… 361
成熟的真核 mRNA 选择性地输出细胞核 …… 362
许多非编码 RNA 也在细胞核内合成和加工 …… 364
核仁是生产核糖体的工厂 …… 366
细胞核含有多种亚细胞核结构 …… 368
小结 …… 372
从 RNA 到蛋白质 …… 372
mRNA 序列编码在一套核苷酸三联体中 …… 372
tRNA 分子将氨基酸与 mRNA 中的密码子匹配 …… 373
tRNA 从细胞核输出前受到共价修饰 …… 375
特异的酶将每个氨基酸与它合适的 tRNA 分子偶联起来 …… 377
RNA 合成酶进行的编辑确保了精确性 …… 378
氨基酸加入到生长多肽链的 C 端 …… 380
RNA 信息在核糖体上解译 …… 381
延伸因子推动翻译向前进行 …… 383
核糖体是一种核酶 …… 386
mRNA 中的核苷酸序列提供了从哪里开始合成蛋白质的信号 …… 388
终止密码子标记翻译的结束 …… 390
蛋白质在多聚体核糖体上合成 …… 391
质量控制机制在翻译的许多阶段发挥作用 …… 393
标准遗传密码中存在较小的变动 …… 393

许多原核生物蛋白质合成的抑制剂是有用的抗生素 …… 396
蛋白质还在合成时就已开始折叠 …… 396
分子伴侣帮助指导许多蛋白质的折叠 …… 397
暴露的疏水区域为蛋白质质量控制提供了关键信号 …… 400
蛋白酶体降解细胞内大部分新合成的蛋白质 …… 401
一种复杂的泛素结合系统标记将要降解的蛋白质 …… 402
许多蛋白质受到可调型破坏的控制 …… 404
异常折叠的蛋白质能够积聚起来，引起破坏性的人类疾病 …… 404
从 DNA 到蛋白质要经过许多步骤 …… 407
小结 …… 408
RNA 世界与生命的起源 …… 408
生命需要自我催化作用（自催化） …… 409
多聚核苷酸不仅能储存信息，还能够催化化学反应 …… 409
在 RNA 世界之前可能存在一个前 RNA 世界 …… 410
单链 RNA 分子能折叠成高度复杂的结构 …… 411
自我复制的分子经历自然选择 …… 412
蛋白质合成是怎样进化的 …… 415
现今所有细胞利用 DNA 作为它们的遗传物质 …… 417
小结 …… 417
7 基因表达的调控 …… 420
基因调控概述 …… 420
多细胞生物中不同类型的细胞含有相同的 DNA …… 421
不同类型的细胞合成不同系列的蛋白质 …… 422
细胞能够改变基因的表达以应答外部信号 …… 422
从 DNA 到 RNA 到蛋白质这条途径中，基因表达可以在许多步骤上受到调控 …… 423
小结 …… 425
基因调节蛋白中的 DNA 结合基序 …… 425
运用细菌遗传学发现了基因调节蛋白 …… 425
DNA 螺旋的外侧可被蛋白质识别 …… 426
DNA 双螺旋的几何学特征由核苷酸序列决定 …… 427
短 DNA 序列是遗传开关的基本组分 …… 427
基因调节蛋白含有能够阅读 DNA 序列的结构基序 …… 429
螺旋-转角-螺旋是最简单也是最常见的 DNA 结合基序之一 …… 430
同源异型结构域蛋白构成一类特殊的螺旋-转角-螺旋蛋白质 …… 432
存在许多种 DNA 结合锌指基序 …… 432
β 折叠也能识别 DNA …… 433
亮氨酸拉链基序不仅介导 DNA 结合，而且介导蛋白质二聚化 …… 435
异源二聚化扩大了基因调节蛋白所能识别的 DNA 序列种类 …… 435
螺旋-环-螺旋也介导二聚化和 DNA 结合 …… 436
还不能预测出所有基因调节蛋白识别的 DNA 序列 …… 438
利用凝胶迁移率变动分析能容易地发现序列特异性 DNA 结合蛋白 …… 439
DNA 亲和层析使序列特异 DNA 结合蛋白的纯化更容易 …… 441
可以确定出一个基因调节蛋白识别的 DNA 序列 …… 442
染色质免疫沉淀技术可以在活细胞中鉴定出基因调节蛋白结合的 DNA 位点 …… 444

小结 …… 444
基因开关如何工作 …… 444
色氨酸阻遏蛋白是在细菌中将基因打开或关闭的一种简单开关 …… 444
转录激活蛋白将基因打开 …… 446
一个转录激活蛋白和一个转录阻遏蛋白调控 lac 操纵子 …… 447
真核细胞中转录调控很复杂 …… 449
真核生物基因调节蛋白从距离调控基因表达 …… 449
真核基因调控区由一个启动子和调节 DNA 序列组成 …… 449
真核基因激活蛋白促进 RNA 聚合酶和通用转录因子在转录起始位点上的组装 …… 451
真核基因激活蛋白修饰局部染色质结构 …… 454
基因激活蛋白协同地发挥作用 …… 456
真核基因阻遏蛋白能以不同的方式抑制转录 …… 456
真核基因调节蛋白常常在 DNA 上装配成复合体 …… 457
调节果蝇发育的复杂基因开关由较小的元件构造而成 …… 459
果蝇 *eve* 基因受到组合调控的调节 …… 461
哺乳动物复杂的基因调控区域也由简单的调节模块构成 …… 463
绝缘子是防止真核基因调节蛋白影响远处基因的 DNA 序列 …… 465
细菌利用可交换的 RNA 聚合酶亚基帮助调节基因转录 …… 466
基因开关是逐渐进化而来 …… 467
小结 …… 468
产生特化细胞类型的分子遗传学机制 …… 468
DNA 重排介导了细菌的相转变 …… 468
一组基因调节蛋白决定了芽殖酵母的细胞类型 …… 470
两种彼此阻遏对方合成的蛋白决定了 λ 噬菌体的遗传状态 …… 472
基因调节环路能用于组建记忆装置以及振荡器 …… 473
昼夜节律钟以基因调节中的反馈环为基础 …… 473
单个蛋白质可以协同一组基因的表达 …… 476
一个关键基因调节蛋白的表达能够引发一整串下游基因的表达 …… 476
真核生物中组合基因调控产生许多不同的细胞类型 …… 478
单个基因调节蛋白能触发整个器官的形成 …… 480
稳定的基因表达模式能够传递到子代细胞 …… 481
染色质结构中广泛的染色体改变可以遗传 …… 482
DNA 甲基化模式在脊椎动物细胞分裂时可以遗传 …… 485
脊椎动物利用 DNA 甲基化将基因锁定在沉默状态 …… 486
基因组印迹需要 DNA 甲基化 …… 487
哺乳动物中 CG 岛大约与 20 000 个基因相连 …… 489
小结 …… 490
转录后调控 …… 491
转录弱化作用引起一些 RNA 分子提前终止 …… 491
选择性 RNA 剪接能从同一个基因产生不同形式的蛋白质 …… 492
自从发现选择性剪接以后，已不得不对基因的定义做出修定 …… 494
在果蝇中性别决定依赖于一系列可调型 RNA 剪接 …… 494
RNA 转录物切割和 poly A 加成的位点发生改变能使一个蛋白质的 C 端发生改变 …… 496
RNA 编辑能改变 RNA 信息的含义 …… 497
从细胞核转运 RNA 可以受到调控 …… 498

有些 mRNA 定位在细胞质中的特定区域 …… 500
结合 mRNA 5′和 3′非翻译区的蛋白质介导了翻译负调控 …… 502
一个起始因子的磷酸化整体上调节了蛋白质合成 …… 503
在翻译起始位点上游 AUG 密码子上的起始能调节真核生物的翻译起始过程 …… 504
内部核糖体进入位点提供了翻译调控的机会 …… 505
mRNA 稳定性的改变能调控基因表达 …… 506
细胞质内 poly A 加成能调控翻译 …… 508
无义介导 mRNA 降解在真核生物中用作一种 mRNA 监视系统 …… 508
细胞利用 RNA 干扰来沉默基因表达 …… 509
小结 …… 510
基因组如何进化 …… 511
拷贝和维持 DNA 的正常机制失灵引起基因组改变 …… 511
两个物种基因组序列的差别与它们独自进化以来经历的时间长短呈比例 …… 512
人类和黑猩猩的染色体非常相似 …… 514
人类和小鼠染色体比较表明了基因组大规模的结构趋异是如何发生的 …… 514
难以重构远古基因组的结构 …… 515
在进化过程中，基因重复和趋异是产生遗传新颖性的关键来源 …… 517
重复的基因发生趋异 …… 518
珠蛋白基因家族的进化表明了 DNA 重复是怎样帮助生物进化的 …… 519
外显子重组可以产生编码新蛋白质的基因 …… 520
基因组序列给科学家留下尚待解开的谜 …… 521
物种的遗传变异为基因组的进化提供了一幅精细的画面 …… 522
小结 …… 523

第 III 部分　方　　法

8　操纵蛋白质、DNA 和 RNA …… 529
细胞分离和培养 …… 530
从组织悬液中分离不同类型的细胞 …… 530
细胞可以在培养皿中培养 …… 532
在无血清，化学成分已知培养基中鉴别特定生长因子 …… 533
真核细胞系是均质细胞的重要来源 …… 534
细胞可以融合形成杂种细胞 …… 535
杂交瘤细胞系是单克隆抗体的持久来源 …… 536
小结 …… 538
细胞组分的分离 …… 538
可以通过超离心分离细胞器和大分子 …… 538
在非细胞体系中可以解释的复杂细胞过程的分子细节 …… 540
可以通过色谱分离蛋白质 …… 541
亲和色谱利用了蛋白质上的特异性结合位点 …… 543
蛋白质的大小和亚基组成可以通过 SDS-聚丙烯酰胺凝胶电泳鉴定 …… 543
超过 1000 种蛋白质可以在一次聚丙烯酰胺双向凝胶电泳中分离 …… 545
选择性切割蛋白质产生不同的多肽片段组合 …… 548
质谱可以用来进行肽段测序和蛋白质鉴定 …… 549
小结 …… 551

分离、克隆 DNA 和 DNA 测序 …… 551
可以用限制性内切核酸酶将大分子 DNA 切成片段 …… 552
凝胶电泳分离不同大小的 DNA 分子 …… 554
纯化的 DNA 分子可在体外用同位素或化学标签进行特异标记 …… 554
核酸杂交反应是检测特异性核苷酸序列的灵敏手段 …… 555
DNA 印迹法、RNA 印迹法和电泳分离的核酸分子的杂交 …… 558
用杂交技术可以在细胞或者染色体上对特定的核酸序列进行定位 …… 560
可以从 DNA 文库中克隆基因 …… 560
两种不同的 DNA 文库用于不同的目的 …… 563
cDNA 克隆含有不间断的编码序列 …… 564
分离的 DNA 片段能够进行快速测序 …… 565
核酸序列用于预测蛋白质的氨基酸序列 …… 567
许多生物的基因组都全部测序完成 …… 568
选择的基因片段可以经聚合酶链反应在试管中克隆 …… 569
细胞内的蛋白质可以利用表达载体进行大量的表达 …… 572
小结 …… 574
分析蛋白质的结构和功能 …… 574
蛋白质晶体的 X 射线衍射可以揭示蛋白质的准确结构 …… 575
分子结构也可以通过核磁共振光谱学进行测定 …… 576
序列相似性可以提供蛋白质功能的线索 …… 577
融合蛋白可以用于分析蛋白质功能和在活细胞中追踪蛋白质 …… 578
亲和层析和免疫沉淀可以测定相互关联的蛋白质 …… 580
蛋白质-蛋白质相互作用可以用双杂交系统进行鉴定 …… 581
噬菌体展示的方法也可以检测蛋白质相互作用 …… 582
蛋白质相互作用可以利用表面等离子体共振来进行实时监测 …… 583
DNA 足迹法可以显示蛋白质在 DNA 分子上结合的位置 …… 584
小结 …… 586
研究基因的表达和功能 …… 586
专题 8-1　经典遗传学回顾 …… 586
经典途径从随机突变开始 …… 589
遗传学筛选鉴别突变体在细胞进程中的缺陷 …… 590
互补测验可以揭示两个突变是否发生在同一个基因内部 …… 591
基因定位可以通过连锁分析 …… 592
寻找同源性可以帮助预测基因的功能 …… 594
报道基因可以揭示基因何时何处表达 …… 594
芯片可以一次监测数千种基因的表达 …… 595
目标突变可以揭示基因功能 …… 597
可以按照要求产生含有突变基因的细胞和动物 …… 598
在细菌和一些低等真核生物中细胞中的正常基因可以直接被改造的突变基因代替 …… 598
改造的基因可以用来在二倍体生物中产生特殊的显性失活突变 …… 599
功能获得突变可为基因在细胞或生物体中发挥的功能提供线索 …… 601
可以重新设计基因来产生有任何想要序列的蛋白质 …… 602
改造的基因可以很容易地插入许多动物的生殖系 …… 602
基因打靶可以产生缺失特异基因的转基因小鼠 …… 602
转基因植物对细胞生物学和农业都是重要的 …… 605

大量收集标记的敲除突变体为检测生物体中每个基因的功能提供了工具 …… 607
小结 …… 608
9 细胞显像 …… 610
显微镜下观察细胞的结构 …… 610
光学显微镜能分辨 0.2μm 的细节部分 …… 613
用相差显微镜和微分干涉相差显微镜清晰地观察活细胞 …… 614
电子技术用于增强和分析图像 …… 615
用于显微镜观察的组织通常要被固定和切片 …… 616
细胞中不同的成分可以被选择性染色 …… 617
通过荧光显微技术对特定分子进行细胞定位 …… 618
抗体可用于检测特定分子 …… 620
使用光学显微镜进行复杂的三维物体成像是可能的 …… 621
共聚焦显微镜通过滤去焦平面外的光获取光学截面 …… 622
电子显微镜分辨细胞的细微结构 …… 624
用于电子显微镜的生物标本需要特别准备 …… 626
特定的大分子能用免疫金电子显微镜进行定位 …… 626
用扫描电子显微镜获取表面图像 …… 628
在透射电子显微镜的高分辨率下通过金属阴影来检测表面特征 …… 629
冰冻断裂和冰冻蚀刻电子显微镜方法获取细胞内部的面层图像 …… 630
负染色和低温电子显微成像术能在高分辨率下观察到大分子 …… 632
多重图像能组合到一起提高分辨率 …… 633
来源于不同方向的图像能被组合到一起进行三维重构 …… 633
小结 …… 634
观察活细胞中的分子 …… 635
使用发光指示剂度量细胞内离子浓度的快速变化 …… 635
几种将不透膜的分子导入细胞的方法 …… 636
“笼状”前体分子的光感应活化作用促进了对细胞内动态的研究 …… 637
绿色荧光蛋白用于标记活细胞或组织中的蛋白质 …… 639
光除了用于对微小物体成像外还可用于操纵微小物体 …… 640
可用放射性同位素标记的分子 …… 641
放射性同位素作为示踪分子用于细胞和组织中 …… 642
小结 …… 643

第 IV 部分　细胞的内部构造

10 膜结构 …… 649
脂双层 …… 659
脂双层的流动性依赖于它的成分 …… 653
质膜包含富含有鞘磷脂、胆固醇和某些膜蛋白的脂质筏 …… 656
脂双层的不对称性在功能上是重要的 …… 656
糖脂在所有质膜表面都有发现 …… 658
小结 …… 659
膜蛋白 …… 659
膜蛋白可以通过多种方法和脂双层联在一起 …… 659
大多数跨膜蛋白借助 α 螺旋构象的多肽链穿过脂双层 …… 661

某些β折叠形成大的跨膜通道 …… 662
许多膜蛋白被糖基化 …… 663
用去污剂能溶解并纯化膜蛋白 …… 665
用血影细胞来研究胞质内侧的质膜蛋白 …… 667
血影蛋白是一个细胞骨架蛋白，通过非共价作用和红细胞膜的胞质面相连 …… 668
血型糖蛋白用一个单个的α螺旋穿过红细胞的脂双层 …… 669
红细胞中的带3蛋白是一个促进阴离子转运的多次跨膜蛋白 …… 671
细菌视紫红质是一个质子泵，通过7个α螺旋穿过脂双层 …… 672
膜蛋白通常形成大复合物来行使功能 …… 673
许多膜蛋白在膜平面中扩散 …… 674
细胞可以将蛋白质和脂分子限定在细胞膜中特定区域 …… 677
细胞表面包被有糖残基 …… 679
小结 …… 680
11 小分子的膜运输和膜的电性质 …… 682
膜运输的原则 …… 683
无蛋白质脂双层对离子是高度不渗透的 …… 683
膜运输蛋白可分为两大类：载体蛋白和通道蛋白 …… 684
主动运输通过与能源偶联的载体蛋白介导 …… 684
离子载体可用作增加膜对特异离子通透性的工具 …… 685
小结 …… 686
载体蛋白和膜的主动运输 …… 686
主动运输可被离子梯度驱动 …… 688
质膜中Na^+驱动的载体蛋白调节胞质溶胶pH …… 690
上皮细胞中载体蛋白的不对称分布是溶质跨细胞运输的基础 …… 690
质膜的Na^+-K^+泵是一种ATPase …… 690
一些Ca^{2+}泵和H^+泵也是P型转运ATPase …… 693
Na^+-K^+泵是维持渗透平衡和稳定细胞体积所必需的 …… 693
专题11-1 胞内水平衡：问题及其解决办法 …… 695
合成ATP的膜结合酶是以相反方向起作用的转运ATPase …… 696
ABC运载体是膜转运蛋白的最大家族 …… 697
小结 …… 699
离子通道和膜的电特性 …… 699
离子通道是离子选择性的并在开关两种状态之间转换 …… 699
动物细胞中的膜电势主要依赖于K^+漏电通道和跨质膜K^+梯度 …… 700
专题11-2 能斯特方程的推导 …… 702
当Na^+-K^+泵停止时，静息电位仅缓慢下降 …… 702
细菌K^+通道的三维结构证明离子通道如何起作用 …… 703
神经细胞的功能依赖于它的纵长结构 …… 705
电压门控阳离子通道在电兴奋细胞中产生动作电位 …… 706
专题11-3 用乌贼巨轴突所做的一些经典实验 …… 709
髓鞘形成增加动作电位在神经细胞中传播的速度和效率 …… 711
膜片钳记录表明单个门通道是以全或无的形式开放 …… 712
电压门控阳离子通道在进化上和结构上是相关的 …… 713
递质门控离子通道在化学突触处将化学信号转换成电信号 …… 713

化学突触可以是兴奋型的或是抑制性的 …… 714
神经肌肉接头处的乙酰胆碱受体是递质门控阳离子通道 …… 715
递质门控离子通道是精神活性药物的主要靶点 …… 717
神经肌肉传递涉及 5 组不同离子通道的相继激活 …… 717
各个神经元都是一个复杂计算装置 …… 718
神经计算至少需要 3 类 K^+ 通道的组合 …… 720
哺乳动物海马中的长程增强效应（LTP）依赖于 Ca^{2+} 通过 NMDA 受体通道的输入 …… 722
小结 …… 724
12 细胞内区域和蛋白质分选 …… 727
细胞内的区域化 …… 727
所有真核细胞具有一套相同的膜包被基本细胞器 …… 727
膜包被的细胞器的形态学关系可以从进化发生的角度得到解释 …… 730
蛋白质可以在各分隔之间以不同的方式运动 …… 732
信号序列和信号斑引导蛋白质至细胞内的正确位置 …… 733
专题 12-1 研究信号序列和蛋白质跨膜运输的方法 …… 735
大部分膜包被细胞器并不是从头形成：细胞器自身提供必需信息 …… 736
小结 …… 736
细胞核和细胞质间的分子运输 …… 736
核孔复合体贯穿核被膜 …… 737
核定位信号引导核内蛋白进入细胞核 …… 738
细胞核输入受体结合核定位信号和核孔蛋白 …… 739
细胞核输出和细胞核输入程序一样，方向相反 …… 740
Ran GTP 酶促使核孔复合体定向转运的发生 …… 741
细胞通过调控进入转运装置的途径实现细胞核和细胞质基质间转运的调节 …… 743
核被膜在有丝分裂过程中分解 …… 744
小结 …… 745
线粒体和叶绿体的蛋白质输入 …… 745
蛋白质易位进入线粒体基质依赖于信号序列和蛋白质易位体 …… 746
线粒体前体蛋白以伸展的多肽链形式输入线粒体 …… 747
线粒体前体蛋白从内膜和外膜的接触点输入线粒体基质 …… 748
ATP 的水解和 H^+ 梯度驱动蛋白质输入线粒体 …… 749
线粒体 hsp70 催化的 ATP 水解重复循环使输入过程得以完成 …… 750
转运插入线粒体内膜和进入膜间隙的蛋白质需要 2 个信号序列 …… 751
引导蛋白质插入叶绿体的类囊体膜需要 2 个信号序列 …… 752
小结 …… 752
过氧化物酶体 …… 754
过氧化物酶体利用氧分子和过氧化氢进行氧化反应 …… 754
一个短的信号序列引导蛋白质输入过氧化物酶体 …… 756
小结 …… 756
内质网 …… 757
膜旁核糖体定义糙面内质网 …… 757
光面内质网在某些特殊细胞中大量存在 …… 758
糙面内质网和光面内质网可以通过离心分离 …… 760
信号序列是在输入糙面内质网的蛋白质中首次发现的 …… 761

一种信号识别蛋白（SRP）引导内质网信号蛋白结合糙面内质网膜上特异性受体 …… 762
多肽链通过易位体的含水孔道 …… 763
内质网膜易位并不总要求多肽链同时进行延长 …… 764
易位后大部分可溶性蛋白质的内质网信号序列被切除 …… 765
在单次跨膜蛋白中，唯一的内质网内在信号序列以跨膜 α 螺旋的形式保留在类脂双层膜中 …… 766
起始转移信号和终止转移信号的组合决定了多次跨膜蛋白的拓扑学结构 …… 767
易位多肽链在糙面内质网腔中折叠组装 …… 770
大多数糙面内质网中合成的蛋白质通过加 *N*-连接寡糖而被糖基化 …… 770
寡糖是标记蛋白质折叠状态的标签 …… 772
未正确折叠的蛋白质被输出内质网在细胞质基质中降解 …… 773
内质网中的错误折叠蛋白质可以引发蛋白质舒展反应 …… 774
一些膜蛋白获得共价连接的糖基磷脂酰肌醇锚点 …… 774
大多数类脂双层膜在内质网中组装 …… 775
磷脂交换蛋白帮助从内质网中将磷脂运送至线粒体和过氧化物酶体 …… 778
小结 …… 778

13 细胞内的膜泡运输 …… 781

膜转运的分子机制及各区室多样性的维持 …… 782
专题 13-1 研究囊泡运输的分子机制的方法 …… 783
存在各种类型的有被囊泡 …… 785
网格蛋白的装配驱动囊泡的形成 …… 785
有被囊泡脱离质膜和去包被都是受调节的过程 …… 787
不是所有的转运囊泡都是球形的 …… 789
单体 GTPase 控制被膜的装配 …… 789
SNARE 蛋白和靶 GTPase 指导膜泡转运 …… 790
相互作用的 SNARE 在重新行使功能前需要被掰开 …… 792
Rab 蛋白帮助确保膜泡锚定的特异性 …… 792
SNARE 可以调节膜融合 …… 794
病毒融合蛋白和 SNARE 可能利用相似的机制起作用 …… 795
小结 …… 796
从内质网到高尔基体的运输 …… 796
蛋白质被 COPⅡ包被的运输泡运出内质网 …… 797
只有正确折叠和装配的蛋白质才能离开内质网 …… 798
从内质网到高尔基体的转运是由管状运输簇介导的 …… 798
内质网回收途径利用分选信号 …… 800
许多蛋白质被选择性地留在它们起功能的区室内 …… 801
高尔基体酶跨膜区域的长度决定了它们的细胞定位 …… 801
高尔基体是由一系列有序的间隔组成的 …… 801
寡糖链在高尔基体中经过加工 …… 803
蛋白聚糖是在高尔基体中装配的 …… 804
糖基化的目的是什么？ …… 805
高尔基体由一系列有序的间隔组成 …… 806
高尔基体中的运输可能是通过囊泡或高尔基体的成熟完成的 …… 807
基质蛋白形成一个帮助高尔基体组织的动态的结构支架 …… 808
小结 …… 809
从高尔基体反面网状结构到溶酶体的运输 …… 809

溶酶体是细胞内消化的主要部位 …… 809
溶酶体是一种异质性细胞器 …… 810
植物和真菌的液泡是多样性的溶酶体 …… 811
物质运输到溶酶体的多条途径 …… 811
6-磷酸甘露糖受体识别反面高尔基体网状结构中的溶酶体蛋白 …… 812
M6P 受体在特异的膜之间穿梭 …… 813
溶酶体水解酶的一个信号区提供 M6P 添加的信号 …… 813
GlcNAc 磷酸转移酶的缺失引起人类一种溶酶体贮积病 …… 814
溶酶体也可能经历胞吐途径 …… 815
小结 …… 815
从质膜进入细胞的物质运输：胞吞作用 …… 815
大的颗粒性物质由特殊的吞噬细胞消化 …… 815
胞吞泡由质膜上的有被小窝形成 …… 816
并非所有的胞饮泡都是网格蛋白包被的 …… 817
细胞通过受体介导的胞吞作用转运细胞外的大分子 …… 818
不从胞内体返回的物质被送到溶酶体 …… 820
有些特异的分子被从早期胞内体运回到质膜 …… 820
在前往晚期溶酶体的途中形成多囊泡体 …… 822
大分子能通过跨细胞作用穿过表皮细胞 …… 823
表皮细胞有两种不同的早期胞内体但是只有一种晚期胞内体 …… 825
小结 …… 825
从高尔基体反面网状结构到胞外的运输：胞吐途径 …… 826
许多蛋白质和脂类被自动地从高尔基体运到细胞表面 …… 826
分泌泡从反面高尔基体网状结构出芽 …… 826
分泌泡中的蛋白质在分泌泡形成的过程中通常要经历蛋白质水解过程 …… 829
分泌泡停在质膜附近等待分泌信号以释放内含物 …… 830
调节型分泌可以是一个质膜区及其下的胞质区的局部反应 …… 830
分泌泡的膜组分被迅速地从质膜上移去 …… 830
极性细胞指导蛋白质从反面高尔基网状结构到达质膜合适的区域 …… 831
指导膜蛋白选择性地运到基底膜的胞质分选信号 …… 832
脂筏可能介导到顶部质膜区的鞘糖脂和 GPI 锚定蛋白的筛选 …… 832
突触泡可直接从内吞泡上形成 …… 834
小结 …… 834
14 能量转换：线粒体和叶绿体 …… 837
线粒体 …… 839
线粒体由内外两层膜和两个内部区隔组成的 …… 841
柠檬酸循环产生高能电子 …… 842
化学渗透过程将氧化能量转化成 ATP …… 842
电子经 3 个大的呼吸过程酶复合体的催化由 NADH 传给 O_2 …… 843
随着电子传递，能量以形成电化学梯度的方式储存 …… 844
质子梯度驱动 ATP 合成 …… 845
质子梯度驱动跨膜转运的机制 …… 846
质子梯度是细胞中 ATP 的主要来源 …… 848
细胞线粒体中 ATP：ADP 比值很高 …… 848

ATP 在细胞中具有举足轻重的作用 …… 848
ATP 合酶还具有水解 ATP 和质子泵的功能 …… 850
小结 …… 851
电子传递链及其质子泵 …… 851
容易移动的质子 …… 851
氧化还原电势决定对电子吸引力的大小 …… 852
专题 14-1　氧化还原电位 …… 853
电子传递释放大量的能量 …… 854
呼吸链中的许多电子载体可用分光光度的方法来识别 …… 854
线粒体内膜上镶嵌有呼吸作用三大酶复合体 …… 856
细胞色素氧化酶的铁铜中心高效催化氧气还原 …… 857
线粒体内膜上电子传递通过随机碰撞调节 …… 859
呼吸作用三大酶复合体氧化还原电势的降低为 H^+ 泵出提供能量 …… 859
原子水平上对 H^+ 泵的机制理解 …… 860
H^+ 载体将电子传递和 ATP 合成过程分开 …… 861
呼吸控制正常情况下抑制链中电子的传递 …… 861
机体自身的解偶联剂将褐色脂肪组织中的线粒体转变成生热器 …… 862
细菌也利用化学渗透机制获取能量 …… 862
小结 …… 863
叶绿体和光合作用 …… 863
叶绿体是质体家族中的一员 …… 864
叶绿体类似线粒体但内含有许多区室化的囊泡 …… 864
叶绿体利用光能进行碳的固定 …… 866
核酮糖二磷酸羧化酶催化碳固定 …… 867
每固定 1 分子 CO_2 需要消耗 3 分子的 ATP 和 2 分子的 NADPH …… 867
某些植物的固碳反应通过区隔化来满足低二氧化碳浓度下的生长 …… 868
光合作用依赖于叶绿素分子的光化学特性 …… 870
一个光合体系是由一个反应中心加一个天线复合体组成的 …… 870
在反应中心，被叶绿素捕获的光能从弱的个体中引出强的给电子体 …… 872
非循环光合磷酸化产生 NADPH 和 ATP …… 872
叶绿体可以通过循环光合磷酸化产生 ATP 而不产生 NADPH …… 874
光系统Ⅰ和光系统Ⅱ有相关结构，与细菌的光合体系相似 …… 874
线粒体与叶绿体中的质子推动力是相同的 …… 875
叶绿体内膜的载体蛋白控制代谢物与胞质的交换 …… 876
叶绿体也执行其他的重要生命物质的合成 …… 877
小结 …… 877
线粒体和质体的遗传系统 …… 877
线粒体和叶绿体含有完整的遗传体系 …… 877
细胞器的生长和分裂决定了细胞中线粒体和质体的数量 …… 879
线粒体和叶绿体基因组的多样性 …… 881
线粒体和叶绿体可能都是从内共生细菌进化来的 …… 881
线粒体基因组的一些特点 …… 882
动物线粒体包括已知的最简单的遗传系统 …… 884
许多细胞器基因含有内含子 …… 884

高等植物的叶绿体含有大约 120 个基因 …… 885
线粒体基因以非孟德尔机制遗传 …… 886
细胞器基因的母系遗传 …… 886
酵母 petite 突变体说明了细胞核在线粒体生物发生中的重要作用 …… 888
线粒体和叶绿体含有由细胞核编码的组织特异性蛋白 …… 889
线粒体的大多数脂类是运入的而叶绿体的大多数脂类是自身合成的 …… 889
为什么线粒体和叶绿体有自己的遗传系统呢? …… 889
小结 …… 891
电子传递链的进化 …… 891
最早的细胞可能通过发酵产生 ATP …… 891
电子传递链使一些厌氧菌能利用非发酵分子作为它们主要的能量来源 …… 891
光合作用的菌体通过不可耗竭的还原力越过主要的进化障碍 …… 892
藻青菌的光合电子传递链产生的大气中的氧和新的生命形式 …… 894
小结 …… 897
15 细胞通讯 …… 901
细胞通讯的普遍原则 …… 901
胞外信号分子与特定受体结合 …… 901
胞外信号分子可以近距离或远距离起作用 …… 903
自分泌信号传递能协调相同细胞群的命运 …… 905
间隙连接使信号转导信息能被周围细胞共享 …… 906
每个细胞被设定程序对特异组合的胞外信号分子做出反应 …… 906
不同细胞能对相同胞外信号分子做出不同的应答 …… 907
只有当分子寿命短时，分子的浓度才能迅速被调节 …… 908
一氧化氮通过与靶细胞内的酶直接结合进行信号传递 …… 909
核受体是配体激活的基因调控蛋白 …… 910
三大类的细胞表面受体蛋白是离子通道偶联受体、G 蛋白偶联受体以及酶联受体 …… 913
大多数激活的细胞表面受体通过小分子和胞内信号蛋白网络传递信号 …… 914
某些胞内信号蛋白起分子开关的作用 …… 916
胞内信号传递复合体增强了反应的速度、效率和特异性 …… 917
模件结合结构域介导胞内信号蛋白之间的相互作用 …… 919
细胞能对逐渐增加浓度的胞外信号做出反应 …… 920
细胞能记忆某些信号的效应 …… 922
细胞能调节其对信号的灵敏性 …… 922
小结 …… 923
G 蛋白偶联的细胞表面受体介导的信号传递 …… 924
三亚基的 G 蛋白解离后传递来自 G 蛋白偶联受体的信号 …… 924
某些 G 蛋白通过调控 cAMP 的生成进行信号传递 …… 927
cAMP 依赖的蛋白激酶（PKA）介导 cAMP 的大多数效应 …… 929
蛋白磷酸酶使得 PKA 和其他蛋白激酶的作用瞬时 …… 930
某些 G 蛋白通过激活磷脂酶 C-β 激活肌醇磷脂信号传递途径 …… 931
Ca^{2+} 充当一个广泛存在的胞内信使 …… 934
Ca^{2+} 振荡的频率影响细胞反应 …… 934
Ca/钙调蛋白依赖的蛋白激酶（CaM-激酶）介导动物细胞中 Ca^{2+} 的许多作用 …… 935
某些 G 蛋白直接调控离子通道 …… 938

嗅觉和视觉依赖于调控环核苷酸门控离子通道的 G 蛋白偶联受体 …… 939
胞外信号扩增通过使用小的胞质调质和酶促级联反应被极大地扩增 …… 942
G 蛋白偶联受体的脱敏依赖于受体的磷酸化 …… 943
小结 …… 944
细胞表面酶联受体介导的信号转导 …… 945
激活的受体酪氨酸激酶自身磷酸化 …… 945
酸磷化的酪氨基酸充当具有 SH2 结构域的蛋白质的停泊位点 …… 949
鸟苷酸交换因子激活 Ras …… 949
Ras 激活一条包括 MAP 激酶的下游丝氨酸/苏氨酸磷酸化级联反应 …… 952
PI3 激酶产生质膜上的肌醇磷脂的停泊位点 …… 954
PI3 激酶/蛋白激酶 B 信号传递途径能促进细胞存活和生长 …… 957
酪氨酸激酶关联受体的活性依赖于胞质酪氨酸激酶 …… 958
细胞因子受体激活 Jak-STAT 信号转导途径，提供了一条快速进入细胞核的途径 …… 959
某些蛋白酪氨酸磷酸酶可以充当细胞表面受体 …… 961
TGF-β 超家族的信号蛋白通过受体丝氨酸/苏氨酸激酶和 Smad 起作用 …… 963
受体鸟苷酸环化酶直接产生 cGMP …… 964
细菌趋化性依赖于由组氨酸激酶关联受体激活的二元系统信号传递途径 …… 966
小结 …… 968
依赖于可调控的蛋白酶解的信号转导途径 …… 969
受体蛋白 Notch 通过切割被激活 …… 969
Wnt 蛋白结合 Frizzled 受体并抑制 β-连环蛋白的降解 …… 970
Hedgehog 蛋白通过彼此作用相反的 Patched 和 Smoothened 的受体复合体起作用 …… 973
多个胁迫和促炎症反应刺激通过 NF-κB 依赖的信号传递途径起作用 …… 975
小结 …… 976
植物中的信号传递 …… 977
多细胞性和细胞通讯在植物和动物中独立进化 …… 977
受体丝氨酸/苏氨酸激酶在植物中行使细胞表面受体功能 …… 978
乙烯激活二元信号传递途径 …… 979
光敏色素检测红光，隐花色素检测蓝光 …… 980
小结 …… 982
16 细胞骨架 …… 984
骨架纤维的自组装和动态结构 …… 984
各种类型的骨架纤维由蛋白质亚单位构成 …… 985
专题 16-1 形成骨架的三种蛋白质纤维 …… 986
多个原纤维形成纤维的组织方式具有优越性 …… 987
成核过程是骨架纤维聚合体形成的限速步骤 …… 987
专题 16-2 肌动蛋白和微管蛋白的聚合 …… 989
微管蛋白和肌动蛋白亚单位“从头到尾”组装形成的骨架纤维具有极性 …… 991
微管和肌动蛋白丝两端的生长速率显著不同 …… 993
纤维的踏车和动态不稳定是微管蛋白和肌动蛋白上绑定的核苷酸水解的结果 …… 994
踏车和动态不稳定需要消耗能量，但却非常有用 …… 997
其他蛋白质聚合体通过偶联核苷酸水解发生的结构改变是细胞运动所必需的 …… 998
微管蛋白和肌动蛋白在真核生物进化过程中显示出高度的保守性 …… 999
成束的、卷曲螺旋的中间纤维 …… 1000

中间纤维赋予动物细胞机械稳定性 …… 1002
药物能够改变纤维的聚合作用 …… 1004
小结 …… 1006
细胞对骨架纤维的调控机制 …… 1006
微管通过一个含有γ微管蛋白的蛋白质复合体成核 …… 1007
动物细胞中，微管源于中心体 …… 1007
肌动蛋白丝通常在质膜处成核 …… 1008
纤维的延长受游离亚单位结合蛋白的限制 …… 1011
结合在纤维旁侧的蛋白质既能稳定它们，也能使它们不稳定 …… 1012
与纤维末端相互作用的蛋白质能够显著改变纤维动力学 …… 1015
细胞中的纤维被构建成高度有序的结构 …… 1016
中间纤维被交联并成束形成强有力的排列 …… 1017
性质不同的交联蛋白构建肌动蛋白丝不同的装配 …… 1018
割断蛋白调控肌动蛋白丝与微管的长度和动力学行为 …… 1022
细胞骨架元件能附着到质膜上 …… 1024
细胞骨架纤维特殊的束形成了横跨质膜的坚固的附着结构：黏着斑、黏着带和桥粒 …… 1025
胞外信号能够诱导主要的细胞骨架纤维重排 …… 1026
小结 …… 1028
分子马达 …… 1028
肌动蛋白为基础的动力蛋白是肌球蛋白超家族的成员 …… 1029
两种类型微管马达蛋白：驱动蛋白和动力蛋白 …… 1032
肌球蛋白和驱动蛋白在结构上的类似暗示了一个共同的进化起源 …… 1033
动力蛋白通过偶联 ATP 水解产生力量致使构象变化 …… 1034
动力蛋白的动力学符合细胞函数 …… 1037
动力蛋白介导的膜被细胞器的胞内运输 …… 1038
马达蛋白的功能可被调控 …… 1040
肌肉收缩依赖于肌球蛋白Ⅱ和肌动蛋白丝的滑动 …… 1042
肌肉收缩由胞质内 Ca^{2+} 浓度的突然升高引起 …… 1044
心肌是设计精确的机器 …… 1046
纤毛和鞭毛是由微管和胞质动力蛋白构成的动力结构 …… 1047
小结 …… 1050
细胞骨架和细胞行为 …… 1050
在酵母中能够容易地分析细胞极化的机制 …… 1050
特殊的 RNA 分子由细胞骨架定位 …… 1053
很多细胞能够爬过固体基底 …… 1053
质膜的外凸由肌动蛋白的聚合驱动 …… 1055
细胞的黏附和收缩允许细胞推进自己前进 …… 1056
外部的信号可以指导细胞迁移的方向 …… 1060
神经细胞复杂的形态特化依靠细胞骨架 …… 1062
小结 …… 1065
17 细胞周期与程序性细胞死亡 …… 1067
细胞周期概述 …… 1068
在所有真核生物中，细胞周期控制系统都是相似的 …… 1070
在酵母中细胞周期控制系统可以进行详细的遗传学研究 …… 1070

在动物胚胎细胞中可以进行细胞周期调控的生化分析 …… 1071
利用培养细胞研究哺乳动物细胞的细胞周期调控 …… 1073
可以通过各种方法研究细胞周期进程 …… 1074
小结 …… 1075
细胞周期控制系统元件 …… 1075
细胞周期控制系统激活细胞周期中的主要进程 …… 1076
控制系统可以在特殊的检测点阻滞细胞周期 …… 1077
检测点一般通过细胞内的负反馈信号调控 …… 1078
细胞周期控制系统的基础是由细胞周期蛋白激活的蛋白质激酶 …… 1078
Cdk 活性可以被抑制磷酸化或者抑制蛋白阻断 …… 1080
细胞周期控制系统依赖于周期性蛋白质水解 …… 1081
细胞周期调控还依赖于转录调控 …… 1081
小结 …… 1082
细胞内的细胞周期事件控制 …… 1083
每个周期中 S 期 cyclin-Cdk 复合体（S-Cdk）起始一次 DNA 合成 …… 1083
M 期 cyclin-Cdk 复合体的激活诱发进入有丝分裂期 …… 1085
进入有丝分裂期会被不完整的 DNA 复制阻断：DNA 复制检测点 …… 1086
M-Cdk 为复制染色体分离做准备 …… 1087
蛋白质酶解诱发姐妹染色单体分离 …… 1087
未结合染色体阻断姐妹染色单体分离：纺锤体结合检测点 …… 1088
有丝分裂期的结束需要 M-Cdk 的失活 …… 1089
G_1 期是 Cdk 稳定失活的一个时期 …… 1090
在哺乳动物 G_1 期细胞中 Rb 蛋白起控制开关作用 …… 1092
细胞周期进程以某种方式与细胞成长保持一致 …… 1093
细胞周期进程被 DNA 损伤阻断以及 p53：DNA 损伤检测点 …… 1094
小结 …… 1097
程序性细胞死亡（细胞凋亡） …… 1097
凋亡由细胞内蛋白水解激酶介导 …… 1098
procaspase 通过与适配蛋白结合激活 …… 1099
Bcl-2 家族蛋白和 IAP 蛋白是细胞死亡程序中的主要细胞内调控因子 …… 1101
小结 …… 1102
胞外的环境控制细胞的分裂、生长和凋亡 …… 1102
有丝分裂刺激细胞分裂 …… 1103
当细胞进入一个特殊的不分裂时期时，细胞会延迟分裂 …… 1103
有丝分裂原激活 G_1-Cdk 和 G_1/S-Cdk …… 1104
不正常的增殖信号促使细胞周期停止或者细胞凋亡 …… 1104
人类细胞有一个内在的控制限制细胞分裂的数量和时间 …… 1106
胞外的生长因子可以刺激细胞生长 …… 1107
胞外的生存因子抑制细胞凋亡 …… 1108
相邻的细胞竞争胞外的信号蛋白 …… 1108
很多类型的动物细胞需要锚定来进行生长和增殖 …… 1109
一些胞外信号蛋白抑制细胞成长、细胞分裂和存活 …… 1111
复杂的细胞分化调控模式产生并维持机体形状 …… 1112
小结 …… 1114

18 细胞分裂的机制 …… 1116
M 期概述 …… 1117
粘连蛋白和凝缩蛋白协助已复制染色体的定形以便分离 …… 1117
细胞骨架机器在有丝分裂和胞质分裂中都发挥作用 …… 1118
两个机制帮助确定有丝分裂总发生在胞质分裂之前 …… 1119
动物细胞的 M 期依赖于发生在之前间期的中心体复制 …… 1120
M 期传统上被分为 6 步 …… 1122
专题 18-1 动物细胞 M 期的主要步骤（有丝分裂和胞质分裂） …… 1122
小结 …… 1126
有丝分裂 …… 1126
微管在 M 期不稳定性显著增强 …… 1127
相反马达蛋白的相互作用和相反极性的微管驱动纺锤体组装 …… 1129
着丝粒把染色体连接到有丝分裂纺锤体上 …… 1131
微管在中期纺锤体中高度活跃 …… 1132
具有功能的双极纺锤体可以在没有中心体的情况下在染色体周围组装 …… 1135
后期会被延迟，直到所有染色体被定位在中期板 …… 1136
姐妹染色单体在后期突然分离 …… 1137
后期 A 的着丝粒微管在两端解聚 …… 1138
推力和拉力都对后期 B 有所贡献 …… 1139
在有丝分裂末期，核被膜在单独的染色体周围重新形成 …… 1141
小结 …… 1141
胞质分裂 …… 1141
有丝分裂纺锤体的微管决定了动物细胞的分裂平面 …… 1142
一些细胞重新配置它们的纺锤体以对称分裂 …… 1143
收缩环中的肌动蛋白和肌球蛋白 II 产生胞质分裂的动力 …… 1145
封闭膜器官在胞质分裂时期必须被分配到子细胞 …… 1147
有丝分裂可以不伴随胞质分裂发生 …… 1147
在高等植物中成膜体指导胞质分裂 …… 1148
高等生物精巧的 M 期是从原生生物的分裂机制逐渐进化而来的 …… 1149
小结 …… 1152

第 V 部分 在“社会背景”中的细胞

19 细胞连接、细胞黏着和细胞外基质 …… 1157
细胞连接 …… 1158
封闭连接通过上皮细胞层形成了一个选择透过性屏障 …… 1159
锚定连接把相邻细胞的骨架相连或者把细胞骨架与细胞外基质相连 …… 1162
黏着连接把细胞间的肌动蛋白束连接在一起 …… 1163
细胞间中间纤维相连的桥粒连接 …… 1163
整联蛋白把细胞与细胞外基质相连形成的锚定连接：黏着斑和半桥粒 …… 1166
间隙连接允许小分子直接在细胞间传递 …… 1167
一个间隙连接由 6 个跨膜连接蛋白亚基组成 …… 1168
间隙连接具有多种功能 …… 1169
间隙连接的通透性可以调节 …… 1170
在植物中，胞间连丝有着和间隙连接一样的功能 …… 1170

小结 …… 1173
细胞间黏着 …… 1173
动物细胞在原地或者迁移之后可以形成组织 …… 1174
解离的脊椎动物细胞可以通过选择性细胞黏着重新装配成组织 …… 1174
钙黏素介导依赖 Ca^{2+} 的细胞黏着 …… 1175
钙黏素在发育中有至关重要的作用 …… 1176
钙黏素通过同源聚合的机制介导细胞黏着 …… 1178
钙黏素通过连锁蛋白与肌动蛋白细胞质骨架相连 …… 1178
在血液中选择素介导短暂的细胞黏着 …… 1180
免疫球蛋白超家族成员介导不依赖 Ca^{2+} 的细胞-细胞黏着 …… 1181
多种类型的细胞表面分子平行发挥作用来介导选择性细胞-细胞黏着 …… 1183
非连接的接触可能起始细胞-细胞黏着，然后连接的接触导向和稳定细胞-细胞黏着 …… 1184
小结 …… 1184
动物的细胞外基质 …… 1185
细胞外基质由其中的细胞形成和定向 …… 1186
黏多糖（GAG）链占了大部分空间并且形成亲水凝胶 …… 1186
一般认为透明质素在组织形态发生和修复中有利于细胞迁移 …… 1187
GAG 链与核心蛋白共价连接形成蛋白聚糖 …… 1188
蛋白聚糖可以调节分泌蛋白的活性 …… 1189
在细胞外基质中，GAG 链可能是高度组织化的 …… 1190
细胞表面的蛋白聚糖作为辅助受体发挥作用 …… 1191
胶原是细胞外基质中主要的蛋白质 …… 1192
胶原分泌时，在每个末端有非螺旋化的延伸 …… 1194
分泌后，纤丝前胶原分子被剪切为胶原分子，然后装配成胶原原纤维 …… 1194
与原纤维结合的胶原有助于原纤维的组织装配 …… 1196
细胞帮助胶原原纤维进行装配，它们通过向基质施加压力分泌产生 …… 1196
弹性蛋白赋予组织以弹性 …… 1197
纤连蛋白是帮助细胞附着在基质上的细胞外蛋白 …… 1199
纤连蛋白可以以纤维状或溶解性状态存在 …… 1200
细胞内肌动蛋白纤维调节细胞外纤连蛋白原纤维的装配 …… 1200
基质中糖蛋白有助于引导细胞迁移 …… 1201
基膜主要由类型Ⅳ胶原、层粘连蛋白、巢蛋白和硫酸类肝素蛋白聚糖组成 …… 1202
基膜具有多种功能 …… 1204
细胞外基质可以影响细胞的形状、生长和增殖 …… 1206
细胞外基质组分的可调控降解有助于细胞迁移 …… 1207
有三种基本的机制来保证蛋白酶对基质组分的降解是受严谨控制的 …… 1208
小结 …… 1209
整联蛋白 …… 1209
整联蛋白是跨膜异源二聚体 …… 1210
整联蛋白必须与细胞质骨架相互作用才使细胞与细胞外基质结合 …… 1211
细胞调节它们整联蛋白的活性 …… 1212
整联蛋白激活细胞内的信号通路 …… 1213
小结 …… 1215
植物细胞壁 …… 1215

细胞壁的成分依赖于细胞类型 …… 1215
植物细胞壁的抗张强度使植物细胞具有膨压 …… 1217
纤维素微原纤维与果胶多聚糖的网状结构交织形成初生细胞壁 …… 1217
微管对细胞壁沉积的导向作用 …… 1219
小结 …… 1222
20 生殖细胞和受精 …… 1225
性别的好处 …… 1225
在多细胞的动物和大多数的植物中，双倍期是长期复杂的，而单倍期是短暂简单的 …… 1226
有性生殖为生活在复杂多变环境里的生物提供了竞争优势 …… 1228
小结 …… 1228
减数分裂 …… 1228
减数分裂中复制的同源染色体配对 …… 1229
配子是经过两次细胞减数分裂产生的 …… 1230
同源非姐妹染色单体之间的交换促进了基因的重排 …… 1230
交叉在减数分裂期染色体分离时有着重要的作用 …… 1233
性染色体的配对确保它们也能彼此分离 …… 1234
联会复合体的形成标志着减数分裂染色体配对达到顶峰 …… 1235
重组结标明了基因重组的位点 …… 1236
遗传图谱揭示了易于发生交换的位点 …… 1236
连续两次不进行 DNA 复制的细胞分裂结束了减数分裂 …… 1237
小结 …… 1239
哺乳动物的原始生殖细胞和性别决定 …… 1239
原始生殖细胞进入正在发育的性腺 …… 1240
Y 染色体上的 *Sry* 基因能让雌性胚胎重新发育成雄性胚胎 …… 1241
小结 …… 1243
卵子 …… 1243
卵子储存了大量的营养物质，而且还有一层精巧的外膜，是高度专一的独立发育 …… 1243
各个时期的卵子发育 …… 1244
卵母细胞采取特殊的机制生长来增长自身的体积 …… 1247
小结 …… 1248
精子 …… 1248
精子高度适应于把它们的 DNA 释放到卵子中 …… 1249
大多数哺乳动物持续地产生精子 …… 1249
小结 …… 1252
受精 …… 1253
物种特异地结合到透明区诱导了精子进行顶体反应 …… 1254
卵子皮层反应确保了仅有一个精子使卵子受精 …… 1255
精子-卵子融合的机制目前仍然不清楚 …… 1256
精子为受精卵提供了一个中心粒 …… 1257
小结 …… 1259
21 多细胞生物的发育 …… 1261
动物发育的通用机制 …… 1261
动物具有相同的基本解剖学特征 …… 1262
多细胞动物调控细胞间相互作用和基因调控蛋白的保守性 …… 1263

调控 DNA 序列决定了发育程序 …… 1264
胚胎学研究揭示了细胞间相互作用机制 …… 1265
通过突变寻找控制发育过程的基因 …… 1266
细胞命运决定早于细胞变化 …… 1267
细胞位置的记忆性影响它们在体内的定位 …… 1267
通过不对称分裂产生两个不同的子代细胞 …… 1268
可诱导的相互作用可以使最初完全一致的细胞产生分化 …… 1269
形态发生素是长距离诱导信号，能够产生多级作用 …… 1270
细胞外信号分子抑制物可以塑造细胞对诱导物的响应 …… 1271
细胞内在程序决定发育的时间程序 …… 1271
最初的图案建立在一小块区域内并随着胚胎生长被顺序诱导 …… 1272
小结 …… 1273
线虫：通过单个细胞来研究发育 …… 1273
线虫解剖学结构简单 …… 1273
线虫发育的细胞命运可以被精确预言 …… 1274
母源基因产物控制了受精卵的不对称分裂 …… 1274
细胞-细胞间相互作用建立了更复杂的图案 …… 1275
显微技术和遗传学揭示了发育控制的逻辑；基因克隆和测序揭示了分子机制 …… 1277
受到发育信号影响后，细胞会发生改变 …… 1277
异时性基因控制了发育的时间表 …… 1277
细胞在计时其内部程序时并不计数细胞分裂 …… 1278
细胞凋亡是发育程序的一部分 …… 1279
小结 …… 1279
果蝇身体图案形成的分子遗传学机制 …… 1279
昆虫身体由一系列体节构成 …… 1280
果蝇从多核体开始其发育 …… 1281
遗传筛选鉴定与特定早期图案建立相关的基因 …… 1283
卵和周围细胞的相互作用决定了胚胎轴：卵极性基因的功能 …… 1284
背腹信号基因建立了核基因调控蛋白的浓度梯度 …… 1286
Dpp 和 Sog 建立了第二个形态发生浓度来精细调控胚胎背部的发育 …… 1286
昆虫的背腹轴对应于脊椎动物的腹背轴 …… 1288
三组体节极性基因决定了前后轴的母源图案并细分胚胎 …… 1288
分节基因的表达受位置信号层次的调控 …… 1289
DNA 调控序列的模块化特性使得基因具有多种独立的功能 …… 1289
卵极性基因、间隙基因和成对控制基因建立了瞬时图案并被其他基因维持 …… 1291
小结 …… 1292
同源选择基因和前后轴的图案形成 …… 1293
HOX 决定了前后的不同 …… 1293
同源选择基因编码了 DNA 结合蛋白并与其他基因调控蛋白相互作用 …… 1293
同源选择基因按照其在 HOX 复合物中的顺序表达 …… 1294
同源异型复合物长期记忆其位置信息 …… 1294
脊椎动物中的同源异型选择基因也控制前后体轴 …… 1296
小结 …… 1298
器官发生和附肢形成 …… 1298

条件诱导的体细胞突变使我们能够研究发育晚期的基因功能 …… 1299
成体果蝇的身体是由成虫盘发育而来 …… 1299
同源异型选择基因对成虫盘的位置信息记忆具有重要作用 …… 1299
特定的调控基因决定将要发育为附肢的细胞 …… 1300
昆虫翅成虫盘的分块现象 …… 1302
4 个重要的信号通路控制翅成虫盘的图案：Wigless、Hedgehog、Dpp 和 Notch …… 1303
每个分区的大小受细胞间相互作用的调控 …… 1303
脊椎动物肢体形成的相似机制 …… 1305
特定类型基因调控蛋白的表达预示了细胞分化 …… 1306
侧向抑制从预神经细胞簇中选出感觉神经母细胞 …… 1307
侧向抑制促使感觉母细胞的后代向最终命运分化 …… 1307
不对称分裂的平面极性由 Frizzled 调控的信号决定 …… 1309
侧向抑制和不对称分裂共同调控身体神经元的产生 …… 1309
Notch 信号调控了不同组织中不同类型细胞构建的美丽图案 …… 1311
一些关键的调控基因决定了细胞类型；其他的激活建立整个器官的程序 …… 1311
小结 …… 1311
细胞运动和脊椎动物体型的形成 …… 1312
胚胎极性依赖于卵极性 …… 1312
细胞分裂由一细胞产生多细胞 …… 1313
原肠胚发育将形成具有简单肠道的三胚层结构 …… 1313
原肠胚运动是可被精确预言的 …… 1315
化学信号引发机械过程 …… 1315
细胞包装改变为原肠胚运动提供了动力 …… 1316
细胞间粘连分子的改变促进细胞重排 …… 1318
脊索延伸，神经板卷曲形成神经管 …… 1318
基因表达振荡器控制了中胚层形成体节 …… 1319
胚胎组织被迁移细胞侵袭 …… 1320
迁移细胞的分布依赖于生存因子和诱导信号 …… 1321
脊椎动物身体的左右不对称性由早期胚胎分子不对称性决定 …… 1321
小结 …… 1323
小鼠 …… 1324
哺乳动物发育具有特异的开始方式 …… 1324
早期哺乳动物胚胎是高度可调控的 …… 1325
从哺乳动物胚胎可以获得全能干细胞 …… 1326
上皮细胞核间充质细胞间相互作用产生了分支管状结构 …… 1327
小结 …… 1328
神经的发育 …… 1328
产生的部位和时间不同导致不同特征的神经元 …… 1329
神经元产生的时间特异性决定了其所形成的相互间连接的特异性 …… 1331
每一个伸展中的轴突和树突顶部都有一个生长锥的结构 …… 1332
在体内，生长锥使神经突起能沿着精确设定的路线生长 …… 1334
生长锥的敏感性在其运动过程中会发生改变 …… 1335
靶组织分泌神经营养因子控制神经细胞的生长和存活 …… 1336
神经元的特异性形成有序的神经元图谱 …… 1337
视网膜不同区域发出的轴突对于视顶盖中抑制分子浓度梯度的响应不同 …… 1339

突触联系模式的重塑取决于突触活动强度 …… 1340
后天经验参与重塑大脑的突触联系模式 …… 1342
成人的记忆机制可能和发育时的突触重塑机制是一致的 …… 1343
小结 …… 1343
植物发育 …… 1343
拟南芥作为一种植物分子遗传学的模式生物 …… 1344
拟南芥基因组富含发育控制基因 …… 1345
胚胎发育开始于根-茎轴向的建立，结束于种子的形成 …… 1346
专题 21-1　开花植物早期发育的特征 …… 1347
专题 21-2　组成高等植物的细胞类型和组织 …… 1348
植物的各部分是由分生组织产生的 …… 1351
幼苗的发育依赖于环境信号 …… 1352
一种新结构的形成依赖于细胞分裂和生长的方向 …… 1352
每一植物模块都是由分生组织的一个原基生长而来的 …… 1353
细胞信号维持着分生区的结构 …… 1354
分生区的突变体可以通过改变细胞的行为来改变植物的拓扑结构 …… 1356
长距离的激素信号在植物的各个部分协同调节植物的发育 …… 1357
同源选择基因特异于花的各个部分 …… 1358
小结 …… 1360
22　组织学：组织中细胞的生存和死亡 …… 1363
表皮及表皮通过干细胞再生 …… 1363
上皮细胞构成一道多层的防水屏障 …… 1364
正在分化的表皮细胞在其成熟过程中合成一系列不同的角质素 …… 1365
表皮由基底层的干细胞更新 …… 1366
干细胞产生的两种子细胞并不一定总是不相同 …… 1366
基底层包含干细胞和短暂扩充细胞 …… 1367
表皮的更新受到许多相互影响的信号控制 …… 1369
乳腺经历发育和退化的重复循环 …… 1370
小结 …… 1371
感觉上皮 …… 1372
嗅觉神经元不断地被更新 …… 1372
听毛细胞的终生持续使用 …… 1373
绝大多数永久细胞进行部分更新：视网膜感光细胞 …… 1375
小结 …… 1376
呼吸道和消化道 …… 1377
不同类型的肺泡旁细胞之间相互协作 …… 1377
杯状细胞、纤毛细胞、巨噬细胞的协作保证了呼吸道的清洁 …… 1378
小肠内壁细胞的自我更新快于其他任何器官 …… 1379
Wnt 信号途径中的成员用于维持消化道干细胞的数量 …… 1381
肝是消化道和血液之间的汇合处 …… 1381
肝细胞的消亡或损失刺激肝细胞增殖 …… 1382
小结 …… 1383
血管和内皮细胞 …… 1383
内皮细胞分布于所有血管内壁 …… 1384

现有内皮细胞的简单分裂产生新的内皮细胞 …… 1385
新的毛细血管通过芽殖形成 …… 1385
血管发生受到外围组织所释放的因子的控制 …… 1387
小结 …… 1388
多能干细胞的更新：血细胞的形成 …… 1388
三种主要的白细胞：粒细胞、单核细胞和淋巴细胞 …… 1389
骨髓中每种血细胞的产生受到分别调控 …… 1391
骨髓中含有造血干细胞 …… 1392
所有类型的血细胞都由多能干细胞产生 …… 1394
血细胞的定型是分步进行的 …… 1394
定型祖代细胞的分裂使特化血细胞的数量扩增 …… 1396
干细胞依赖基质细胞发出的接触信号 …… 1396
调节造血的因子可以通过体外培养进行研究 …… 1397
红细胞的生成依赖促红细胞生成素 …… 1397
多种集落刺激因子（CSF）影响中性粒细胞和巨噬细胞的产生 …… 1398
造血细胞的行为具有一定的随机性 …… 1399
细胞存活调节和细胞增殖调节同等重要 …… 1400
小结 …… 1400
骨骼肌的发生、调整和再生 …… 1401
新的骨骼肌纤维由肌原细胞融合产生 …… 1401
肌细胞通过改变胞内蛋白质异型体改变自身特征 …… 1403
骨骼肌纤维分泌筒箭毒碱限制自身的生长 …… 1403
一些肌原细胞以静息干细胞的形式存在于成年个体内 …… 1404
小结 …… 1405
成纤维细胞及其转化：结缔组织细胞家族 …… 1405
成纤维细胞受到不同化学信号的刺激改变自身的性质 …… 1405
细胞外基质可能通过影响细胞形状及其附着来影响结缔组织细胞的分化 …… 1407
脂肪细胞可以从成纤维细胞发育而来 …… 1407
脂肪细胞分泌的瘦素蛋白通过负反馈抑制进食活动 …… 1408
骨骼由其细胞不断改造重建 …… 1409
成骨细胞分泌骨基质，而破骨细胞则侵蚀之 …… 1410
发育过程中，破骨细胞侵蚀软骨为骨骼的发生创造空间 …… 1412
小结 …… 1413
干细胞工程 …… 1414
胚胎干细胞可用于产生身体的任何部分 …… 1414
表皮干细胞可在体外扩增培养用于组织修复 …… 1416
神经干细胞可用于补充中枢神经系统细胞 …… 1416
成年个体组织的干细胞可能比想象的更具多能性 …… 1417
小结 …… 1417
23 癌症 …… 1419
癌症的形成是一个微进化过程 …… 1419
癌细胞的无限增殖和移居异质组织 …… 1420
大多数癌症来源于单个异常细胞 …… 1422
由体细胞突变引起的癌症 …… 1422

单个突变不足以引发癌症 …… 1423
从轻度变异细胞缓慢发展而来的癌症 …… 1424
进行性肿瘤包含连续几轮突变和自然选择 …… 1426
大多数人类癌细胞在遗传上是不稳定的 …… 1427
癌性生长常常依赖于对细胞死亡和细胞分化控制的缺陷 …… 1428
许多癌细胞都脱离固有的细胞增殖界限 …… 1430
恶性癌细胞在一种异己环境中生存和增殖才能转移 …… 1431
6 个关键特性使细胞能够进行恶性生长 …… 1431
小结 …… 1433
可预防的癌症病因 …… 1433
许多但不是所有的致癌因子都损伤 DNA …… 1434
不改变细胞 DNA 序列的因素能加速癌瘤发生 …… 1435
病毒和其他感染促成一部分重要的人类癌症 …… 1437
鉴别致癌物展现出避免癌症的途径 …… 1438
小结 …… 1439
癌症关键基因的发现 …… 1440
鉴别功能获得型突变和功能缺失型突变的各种方法 …… 1440
癌基因通过其显性转化效应而被鉴定 …… 1440
肿瘤抑制基因有时可通过研究罕见的癌遗传综合征来鉴定 …… 1442
即使没有从癌遗传综合征获得线索也能鉴定肿瘤抑制基因 …… 1443
癌瘤中的突变基因可通过多种途径而激活过度或不足 …… 1444
寻找肿瘤关键基因仍在继续 …… 1446
小结 …… 1446
癌细胞行为的分子基础 …… 1447
研究发育中的胚胎和转基因小鼠有助于揭示癌症关键基因的功能 …… 1447
许多癌症关键基因调节细胞分裂 …… 1449
调控细胞凋亡的基因的突变允许癌细胞逃离自杀 …… 1450
p53 基因的突变允许癌细胞在发生 DNA 损伤时存活并增殖 …… 1451
DNA 肿瘤病毒通过阻断关键肿瘤抑制基因的行为激活细胞复制装置 …… 1452
端粒缩短可能为人类癌症铺路 …… 1453
在一群端粒缺陷的细胞中，p53 的丢失打开了一条通往癌症的林荫大道 …… 1454
导致转移的突变仍然是一个谜 …… 1455
结肠直肠癌通过一连串的可见突变而缓慢演化 …… 1457
大多数结肠直肠癌病例的共同点是都有几个关键的遗传损伤 …… 1459
DNA 错配修复的缺陷是患结肠直肠癌的另一途径 …… 1460
进行性肿瘤的每一步可能都与特异突变相关 …… 1461
每例癌症都有一组特异的遗传损伤 …… 1462
小结 …… 1462
癌症治疗：现在与未来 …… 1463
寻找癌症疗法是困难的，但并非没有希望 …… 1463
当前疗法利用细胞周期控制的丧失和癌细胞的遗传不稳定性 …… 1463
癌症能产生对治疗的抗性 …… 1463
新疗法可能会出自我们的癌症生物学知识 …… 1464
可将治疗设计成攻击缺乏 p53 的细胞 …… 1465

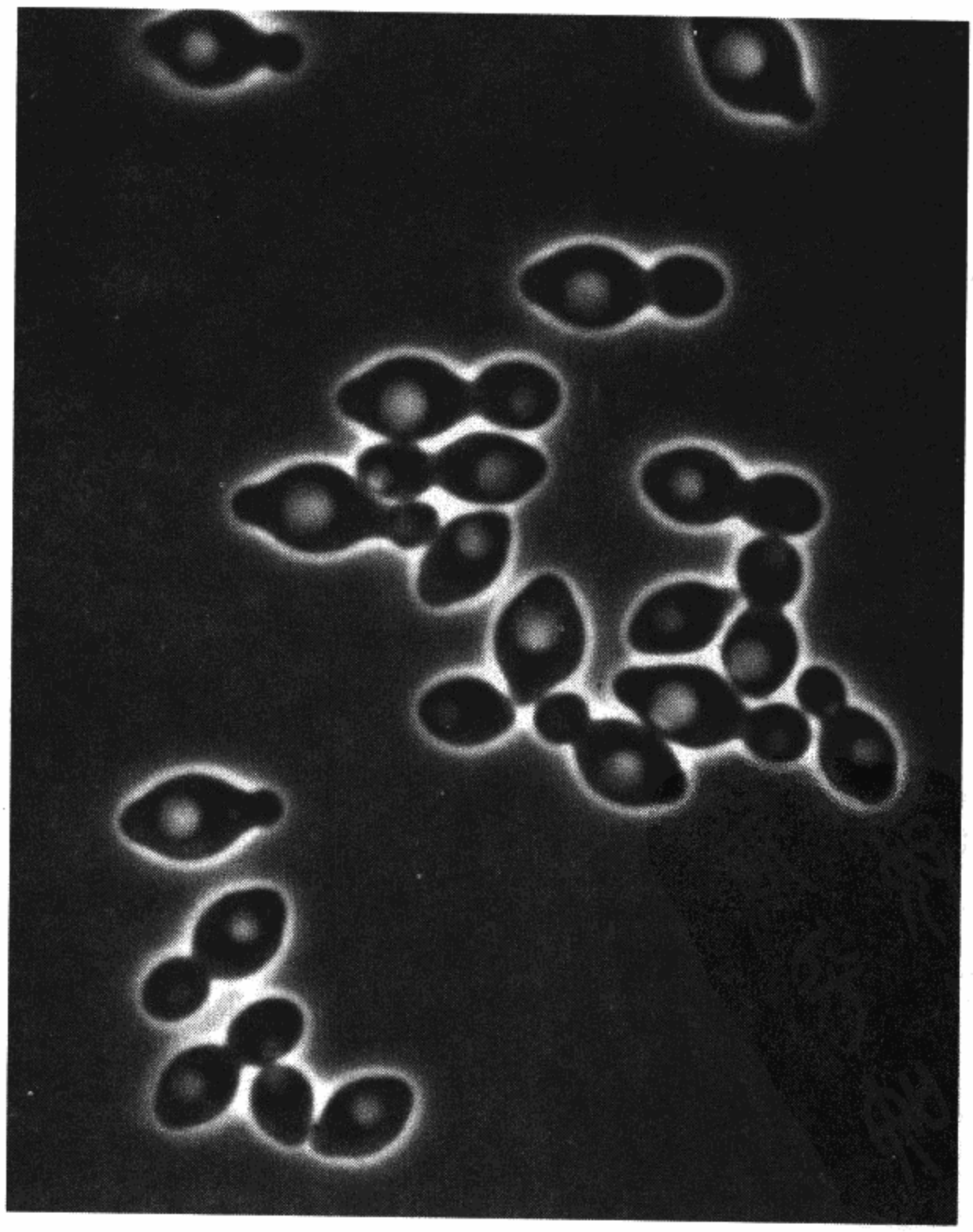
肿瘤生长可通过切断癌细胞的血液供应而被阻断 …… 1465
可以设计以特异癌基因蛋白为靶的小分子 …… 1465
对细胞生物学的理解导致合理而简明的内科治疗 …… 1466
小结 …… 1468
24 获得性免疫系统 …… 1470
淋巴细胞和获得性免疫的细胞基础 …… 1471
获得性免疫需要淋巴细胞 …… 1472
天然免疫系统和获得性免疫系统一起发挥作用 …… 1472
B 淋巴细胞在骨髓中发育；T 淋巴细胞在胸腺中发育 …… 1474
获得性免疫系统通过克隆选择发挥作用 …… 1475
大多数抗原激活许多不同的淋巴细胞克隆 …… 1477
免疫记忆是由于克隆扩增和淋巴细胞分化的结果 …… 1477
获得性免疫耐受确保自身抗原免受攻击 …… 1478
淋巴细胞不断地循环通过外周淋巴器官 …… 1480
小结 …… 1482
B 细胞和抗体 …… 1483
B 细胞使抗体成为细胞表面受体和分泌分子 …… 1483
一个典型的抗体有两个完全相同的抗原结合位点 …… 1484
抗体分子由重链和轻链组成 …… 1485
重链可分为 5 大类，每类都具不同的生物学特性 …… 1485
抗原抗体相互作用强度依赖于抗原结合位点的数目和亲和力 …… 1489
轻链和重链由恒定区和可变区组成 …… 1491
轻链和重链由重复的 Ig 结构域组成 …… 1492
抗原结合位点由超变环构成 …… 1492
小结 …… 1494
抗体多样性的产生 …… 1494
抗体基因是在 B 细胞发育过程中由独立的基因片段组装而成 …… 1495
每个可变区都由一个以上的基因片段编码 …… 1496
基因片段的不精确连接大大增加了 V 区的多样性 …… 1496
抗原驱动的体细胞突变精确调整抗体反应 …… 1498
V(*D*)*J* 连接的控制保证了 B 细胞的单特异性 …… 1499
被抗原激活后，B 细胞从制造膜结合抗体转变成制造分泌型的相同抗体 …… 1500
B 细胞能转换它们所制造的抗体的类 …… 1500
小结 …… 1501
T 细胞和 MHC 蛋白 …… 1502
T 细胞受体是类抗体异二聚体 …… 1502
抗原呈递细胞激活 T 细胞 …… 1503
细胞毒效应 T 细胞诱导受感染的靶细胞自杀 …… 1504
辅助效应 T 细胞帮助激活巨噬细胞、B 细胞和细胞毒 T 细胞 …… 1506
T 细胞识别结合于 MHC 蛋白的外来多肽 …… 1507
在了解其功能之前，MHC 蛋白是在移植反应中被鉴定出来的 …… 1507
MHC Ⅰ类和 MHC Ⅱ类蛋白是结构上类似的异二聚体 …… 1508
MHC 蛋白与多肽结合并与 T 细胞受体相互作用 …… 1510
MHC 蛋白帮助指挥 T 细胞到达其适当的靶目标 …… 1513

CD4 和 CD8 共受体结合至 MHC 蛋白的不变部分 …… 1513
细胞毒 T 细胞识别与 MHC Ⅰ类蛋白结合的外来胞质溶胶蛋白片段 …… 1514
辅助 T 细胞识别与 MHC Ⅱ类蛋白有关的内吞外来蛋白质的片段 …… 1516
可能有用的 T 细胞在胸腺中受到正选择 …… 1518
能被自身肽活化的许多发育中的 T 细胞在胸腺中被剔除 …… 1518
MHC 蛋白的功能解释其多态性 …… 1520
小结 …… 1520
辅助 T 细胞和淋巴细胞的激活 …… 1521
抗原呈递细胞表面的共刺激蛋白帮助激活 T 细胞 …… 1521
辅助效应 T 细胞的亚类决定获得性免疫应答的性质 …… 1524
T_H1 细胞在感染位点帮助激活巨噬细胞 …… 1526
抗原结合给 B 细胞提供第一个信号 …… 1527
辅助 T 细胞给 B 细胞提供第二个信号 …… 1529
免疫识别分子属于一个古老的超家族 …… 1530
小结 …… 1532
25 病原体、感染和先天性免疫 …… 1535
病原体导论 …… 1535
病原体进化出与其宿主相互作用的特殊机制 …… 1536
感染的体征和症状可能由病原体或宿主的应答而引起 …… 1537
病原体在种系发生上具有多样性 …… 1537
细菌病原体携带专门的致病基因 …… 1538
真菌和原生动物寄生虫有多种形式的复杂生活史 …… 1542
病毒从其增殖的各个角度利用宿主细胞的机制 …… 1544
朊病毒是传染性蛋白质 …… 1547
小结 …… 1548
感染的细胞生物学 …… 1548
病原体跨过保护性屏障在宿主内定居 …… 1548
定居于上皮细胞的病原体必须避免被宿主清除 …… 1549
细胞内病原体进入和离开宿主细胞的机制 …… 1551
病毒与展示在宿主细胞表面的分子结合 …… 1552
病毒通过膜融合、孔形成或膜破裂进入宿主细胞 …… 1553
细菌通过吞噬作用进入宿主细胞 …… 1555
细胞内寄生物积极侵入宿主细胞 …… 1556
许多病原体改变宿主细胞内的膜运输 …… 1558
病毒和细菌利用宿主细胞的细胞骨架在胞内运动 …… 1561
病毒接替宿主细胞的代谢作用 …… 1563
病原体能改变宿主有机体的行为以便于病原体的传播 …… 1565
病原体的快速进化 …… 1565
耐药性病原体是一个日趋紧张的问题 …… 1568
小结 …… 1568
天然免疫 …… 1569
上皮细胞表面有助于预防感染 …… 1569
人的细胞识别病原体的保守特征 …… 1570
补体激活将病原体作为吞噬作用或溶菌作用的靶目标 …… 1571

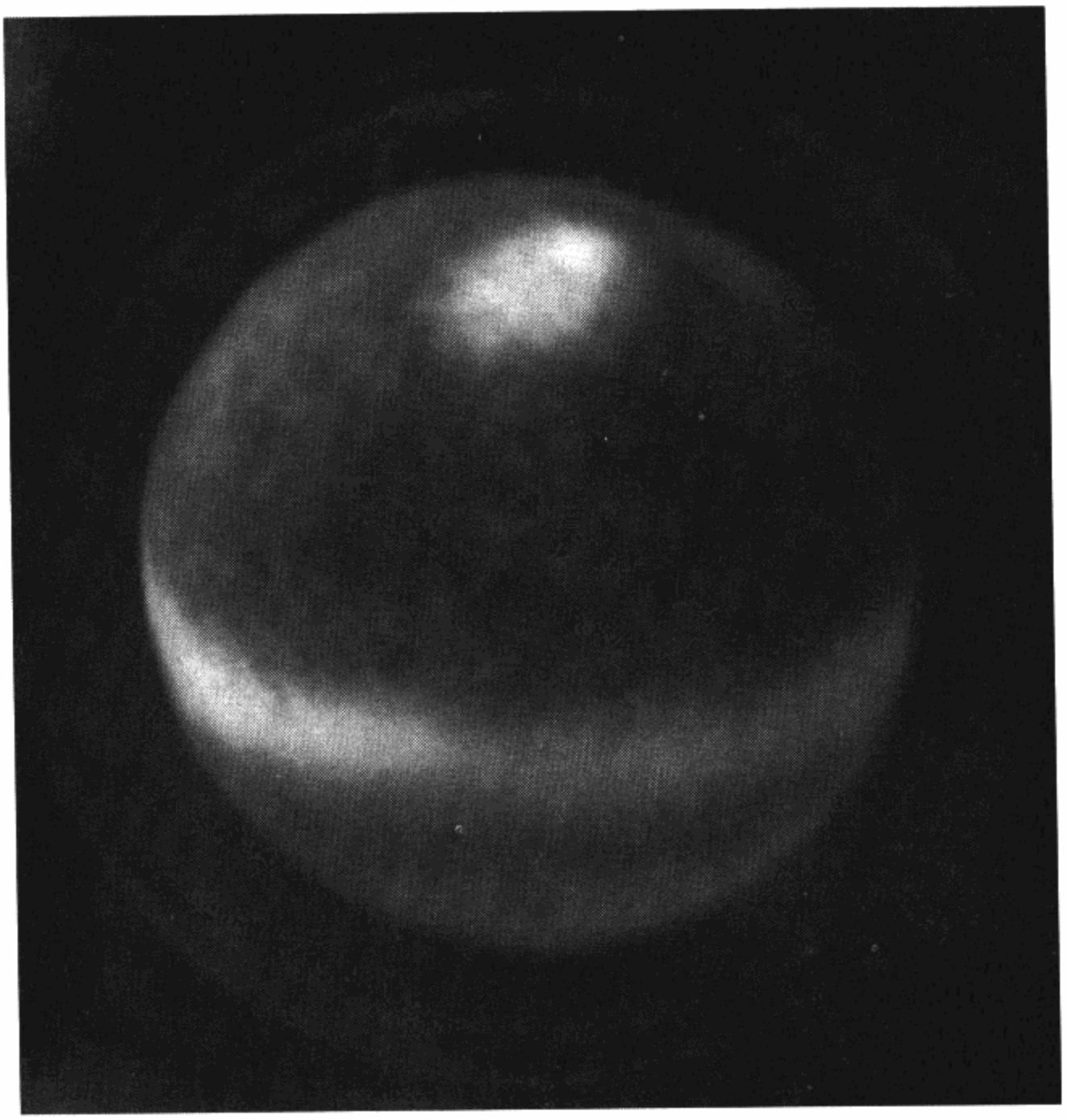
Toll-like 蛋白是模式识别受体的一个古老家族 …… 1573
吞噬细胞寻找、吞噬和消灭病原体 …… 1575
活化的巨噬细胞招募其他吞噬细胞至感染位点 …… 1576
病毒感染细胞采取有力措施阻止病毒复制 …… 1577
天然杀伤细胞诱导病毒感染细胞自杀 …… 1578
小结 …… 1579

第Ⅰ部分　细胞导论

1　细胞和基因组
2　细胞的化学和生物合成
3　蛋白质



分裂中的酵母细胞。这些单细胞有机体通过出芽形成子细胞而迅速分裂。酵母作为一种模式生物被广泛利用，特别是在研究细胞周期中起到重要作用



卵细胞。这单个细胞中含有决定一个多细胞动物结构所必须的遗传信息。这是一个蟾蜍 *Xenopus laevis* 卵细胞

1　细胞和基因组

地球上细胞的共同特征
基因组的多样性及生命树
真核生物的遗传信息

我们的星球表面布满了活体生物——这些古怪的、杂乱的但具有组织构造的“化学工厂”从周边的环境中摄取物质并利用这些原材料制造出千千万万个自身的复制体。这些活体通过各种各样完全不同的方式产生。然而，在老虎与海藻之间，或在细菌与树木之间，除了叫法不同还有没有更多其他差异的地方呢？虽然我们的祖先并不了解什么是细胞和 DNA，但他们发现所有的这些活体都有个共同的特点，他们把它叫做“生命”。人们惊异于它的存在，努力的想要了解它，拼尽全力的想要说明生命是什么，它是怎么区别于非生命体运行的。

20 世纪的重大发现并没有减少人们对于生命的好奇——反而更加引起了人们的兴趣。但这些发现却揭去了自然界生命的神秘面纱。我们现在明白所有的生命体都是由细胞组成，这些生命体的微小单元的大多数基本功能都采取相同的运行模式。生物体虽然外表千差万别，但本质上却大同小异。整个生物学就是两种主题的对应点：一种是个体的多样性，另一种是基本机制的稳定性。在这一章里，我们首先要描述所有生命体的共同特征，接着简要阐明细胞的多样性。我们将会看到，由于所有生命体的规范都表现为共同的密码，所以有可能解读、测量和破译这些规范以便清晰连贯地理解各种形式的生命体。

地球上细胞的共同特征

现在，估计有 1000 万以上——也许有 1 亿物种生活在我们的地球上。每个物种都不相同，它们忠实的复制自己，产生出属于本物种的后代：母体以一种极其详尽的方式将特定的遗传信息传递给子代，使之具有本物种的特征。这种遗传现象是生命这个定义的中心部分，它把生命和其他过程相区分。例如，晶体的形成、蜡烛的燃烧、波浪以及水的生成等，这些过程里亦有有序结构的形成，但没有亲代与子代的特性之间相同类型的联系。和蜡烛燃烧一样，活体必须消耗自由能以产生并维持自身的机体组成；但这些自由能是在遗传信息的指定下驱动一个庞大的化学反应复合系统进行工作的。

大多数生命体都是单细胞结构；其他生命体，像我们一样，是庞大的多细胞都市，体内成簇的细胞执行特定功能，它们之间通过复杂的通讯系统互相连接。但是不管是单细胞细菌还是由 10^{13} 个细胞组成的人体，其所有的组织器官都是由于单个细胞的分裂而来。这单个细胞承载着决定物种的遗传信息（图 1-1）。在这些信息的指引下，细胞从环境中聚集原材料，依照自己的影像在体外复制一个同样拥有全套拷贝的新生细胞。只有细胞有这个能力。

所有的细胞都以同样的线性化学密码（DNA）形式储存遗传信息

我们知道，计算机可以将信息作为一种可测量的量——一张软盘上有 100 万个字节（相

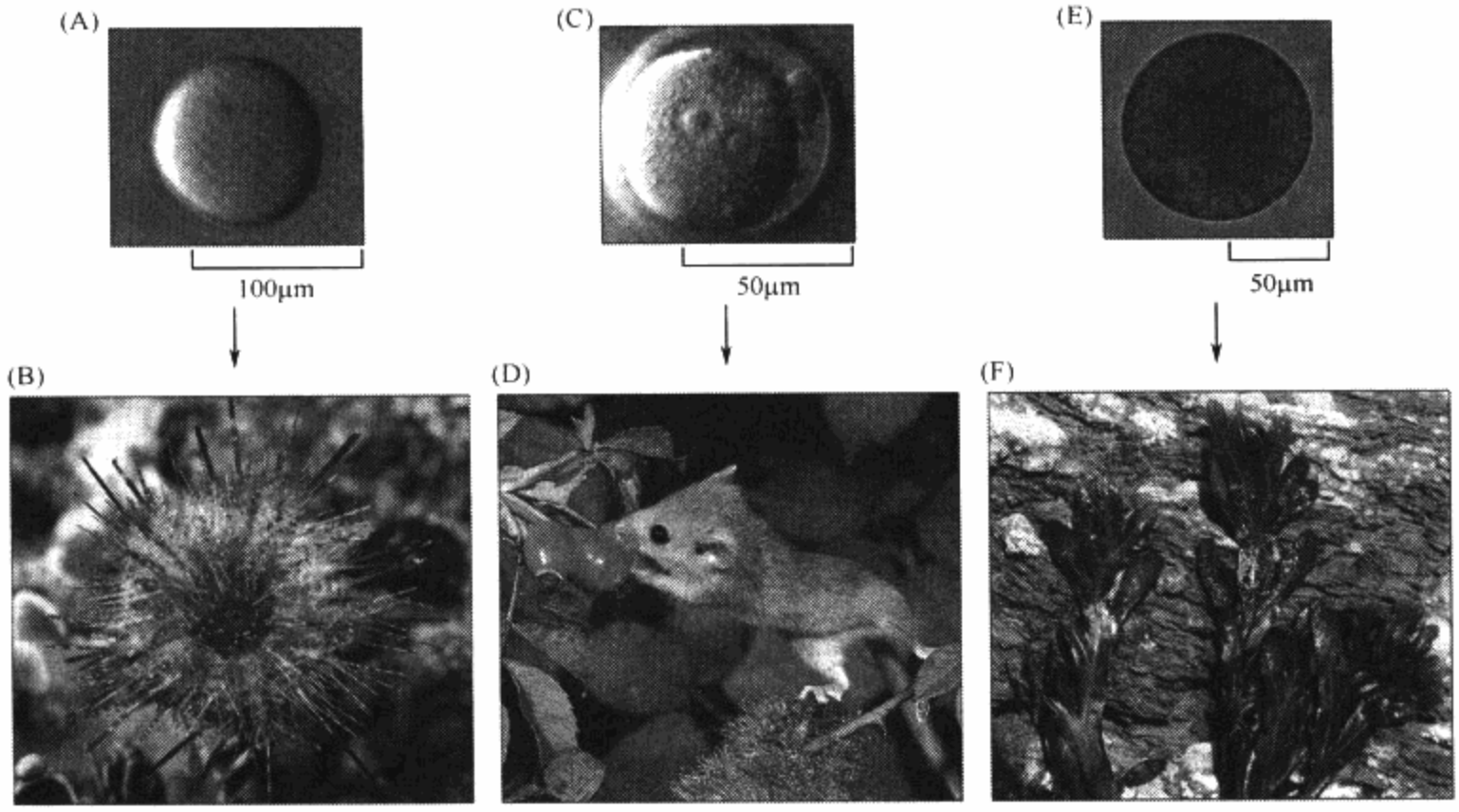


图 1-1　卵细胞中的遗传信息决定了整个多细胞生物体的属性

（A）和（B）海胆卵细胞产生了海胆，（C）和（D）小鼠卵细胞发育成小鼠，（E）和（F）海藻 *Fucus* 的卵细胞形成海藻（A，Courtesy of David McClay；B，Courtesy of M. Gibbs，Oxford Scientific Films；C，Courtesy of Patricia Calarco，from G. Martin，*Science* 209：768-776，1980；D，Courtesy of O. Newman，Oxford Scientific Films；E and F，Courtesy of Colin Brownlee.）

当于 200 页的文本），一张 CD 光盘上可存储 6 亿个字节等。但由于一种信息可以以不同的物理形式存储，从而给我们的生活带来许多不便。在一种型号的计算机上输入的文档可能无法在另一台机器上读取。随着计算机技术的发展，今天的机器可能无法读取 10 年前我们记录电子信息的磁盘和磁带。活细胞和计算机一样也要交换信息，并且它们已经进化发展了 35 亿年。没有人认为它们会以一种相同的形式存储信息，不然一种类型细胞的信息会被其他物种细胞的信息处理器所读取。但是出乎人们的意料，世界上所有的活细胞，除了目前未发现的以外，都无一例外的采用一种双链分子的形式存储遗传信息，这种分子叫做 DNA——没有分支的聚合长双链，永远只由 4 种单体组成：A、T、C、G。这些单体紧密地连在一起，组成一条线性的长序列，编码遗传信息，就像计算机语言中的 1 和 0 一样。我们可以从人细胞中取出一段 DNA 插入细菌中，也可从细菌 DNA 中取出一段插入人类细胞中，得到的信息可以成功地通读、翻译、复制。科学家们利用化学方法可以通读任何 DNA 分子上的全部单体（即使有上百万个碱基）从而可以译解每个生物体的遗传信息。

细胞通过依照模板的聚合作用复制遗传信息

如果要了解生命运行的机制，必须清楚双链 DNA 分子的结构。DNA 单链上的每一个单体，即每一个**核苷酸**都由两部分组成：一个磷酸基团的糖分子（脱氧核糖）和一个碱基。碱基可以是腺嘌呤（A）、鸟嘌呤（G）、胞嘧啶（C）和胸腺嘧啶（T）中的一种（图 1-2）。糖分子之间通过磷酸基团相互连接，这种重复的糖-磷酸结构形成聚合链的骨架，并从骨架上伸出一系列碱基。DNA 聚合体通过将单体添加在它的末端而得以延伸。从理论上来说，一条单链上单体的添加方式是随意的，因为每个单体都以相同的方式与下一个单体相连。然而在细胞中却有这样的限制：DNA 分子并不能随意合成，而是以预先存在的 DNA 链为模板进行合成。由于碱

基间的互补结构：A 与 T 结合，C 与 G 结合，使得合成链上的碱基紧密结合于模板链上。这种碱基配对决定 4 种碱基中的哪种应该连到正在延伸的链上。按照这种方式，就产生一个含有 A、C、T、G 4 种碱基的互补双链，两条链之间互相缠绕，形成一条双螺旋［图 1-2（E）］。

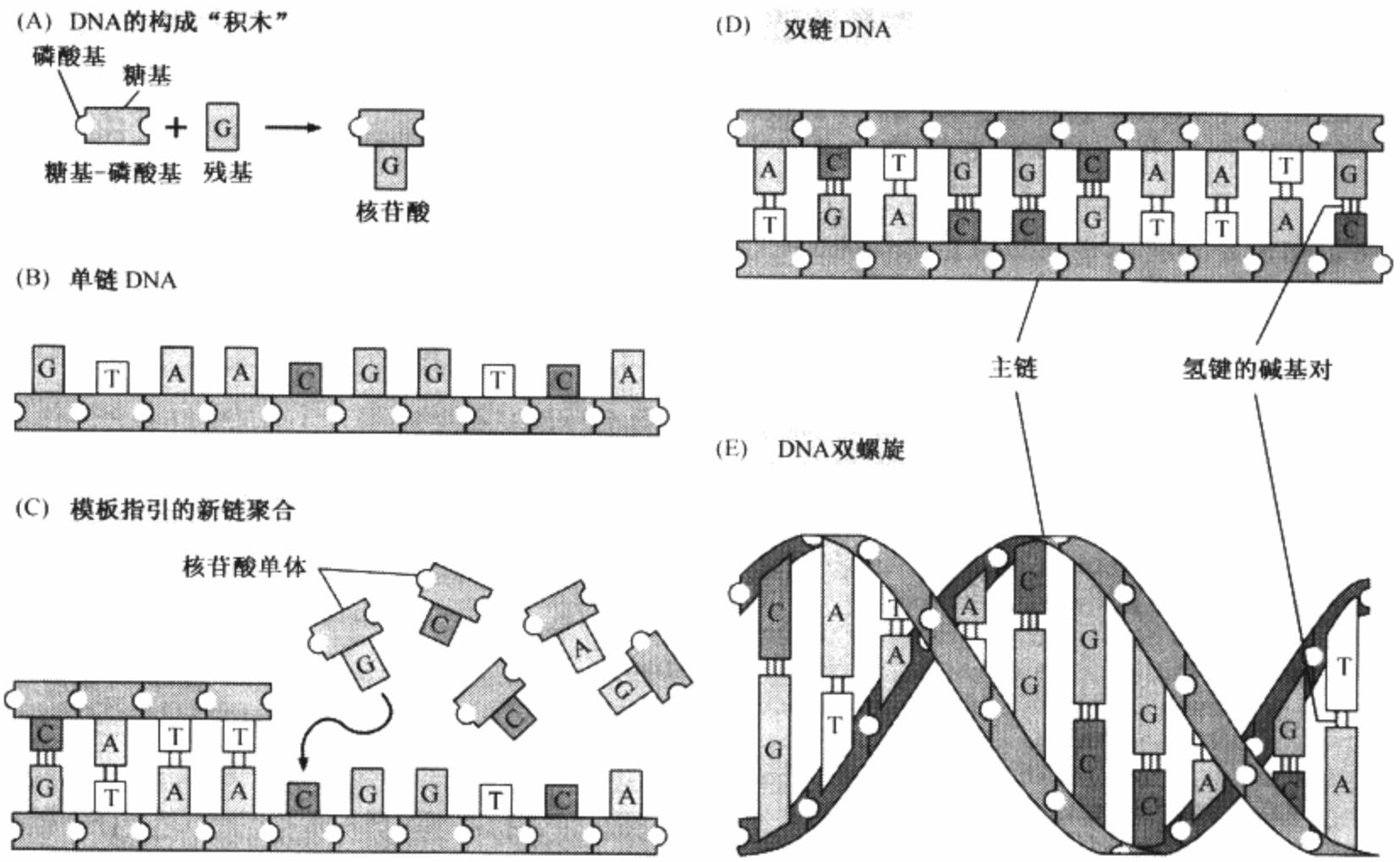


图 1-2　DNA 及其构造单元

（A）DNA 由称为核苷酸的单体组成，每一个核苷酸都由两部分组成：含有一个磷酸基团的糖分子（脱氧核糖）及一个碱基，碱基有 4 种类型（腺嘌呤、鸟嘌呤、胞嘧啶和胸腺嘧啶），分别对应为 A、G、C 和 T。（B）单链 DNA 由核苷酸糖基-磷酸基连接组成。注意单个糖基-磷酸基单元是不对称的，导致了单链骨架的方向性或极性。（C）通过依赖模板的聚合作用，当前 DNA 链指导核苷酸聚合形成新的 DNA 链；T 和另一条链上的 A 结合，G 和 C 结合。新链的核苷酸序列和旧链互补，骨架有着相反的方向性：如新链上的…TTAC 对应于旧链的 GTAA…（D）一般，DNA 分子含有两条这样的互补链。每条链上的核苷酸通过强（共价）化学键相连，而互补两条链间的结合力较弱，通过氢键相连。（E）两条 DNA 链互相缠绕形成双螺旋，这种结构非常牢固，可以在不破坏核苷酸基本结构的情况下调节核苷酸序列

与糖-磷酸连接相比，碱基对之间的键较弱，这允许两条 DNA 链可以在单链骨架不被破坏的情况下被分开。每条链都可以按照上述方式合成一条和自身完全互补的新 DNA 链——遗传信息的新拷贝（图 1-3）。在不同的细胞中，**DNA 复制**（DNA replication）的比例不同，起始结束的时间不同，辅助分子也不同。但是其基本要素是相同的：DNA 是遗传信息的储存库，并且有模板的聚合作用（templated polymerization）是整个生物界复制信息的方式。

所有的细胞都将其部分遗传信息转录成共同的中间体（RNA）

为了行使储存遗传信息的功能，DNA 不仅仅是必须在每次细胞分裂前以上述方式完成自身的复制，它还必须表达自身的信息，用于指导细胞内其他分子的合成。这些过程在所有的生物体中采取相同的机制，产生另外两种重要的聚合体：RNA 和蛋白质。首先是一个依赖于

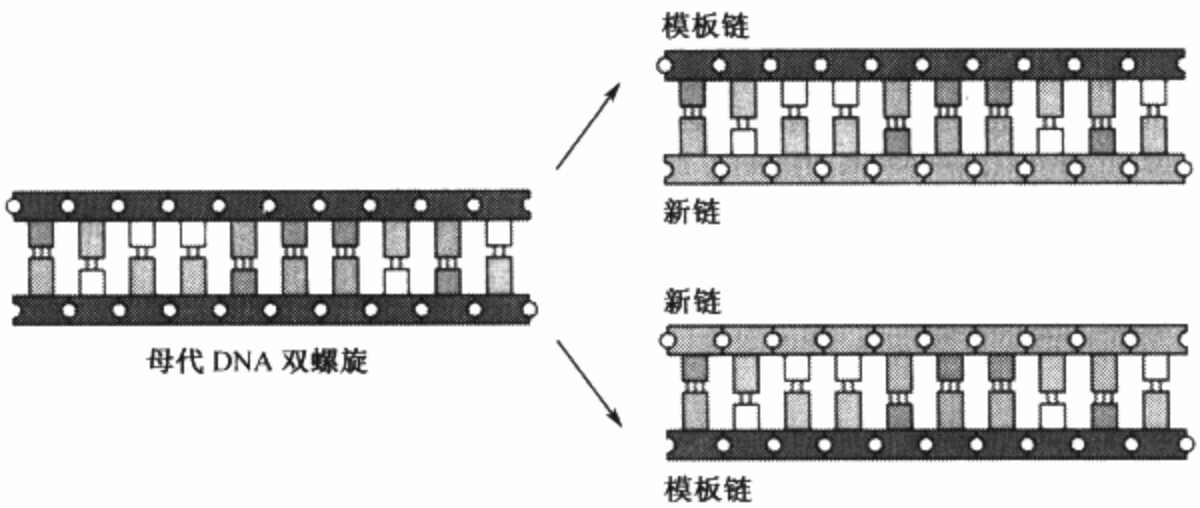


图 1-3　遗传信息通过 DNA 复制而加倍

在此过程中，双链 DNA 解旋，每个 DNA 单链都可以作为模板合成一条互补链

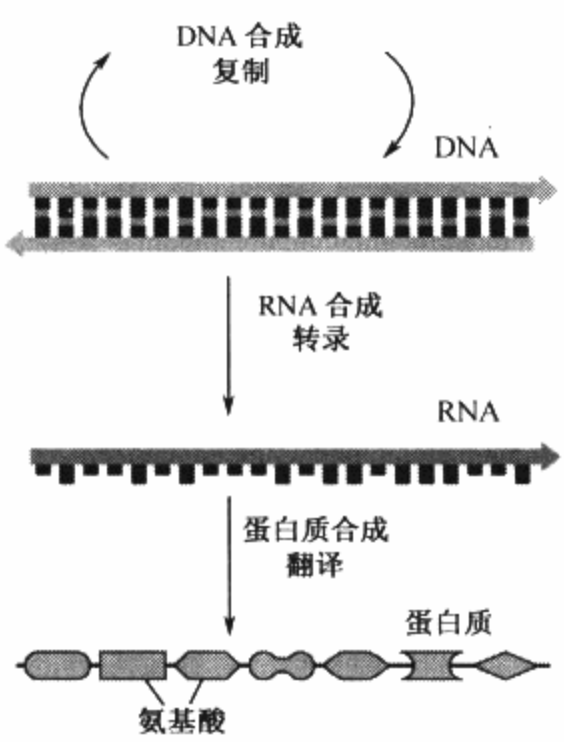


图 1-4　从 DNA 到蛋白质

遗传信息通过两个步骤而被通读和应用。首先，在转录中，DNA 片段指导 RNA 分子的合成。然后在翻译中，RNA 分子指导蛋白质分子的合成

模板的聚合作用，叫做**转录**（transcription），即以 DNA 片段为模板指导合成与之密切相关的小分子聚合物**核糖核酸**（ribonucleic acid）或 **RNA**。然后是一个更加复杂的过程，叫做**翻译**（translation），其中许多 RNA 分子参与这一过程，指导合成另一种完全不同类型的化合物——蛋白质（protein）（图 1-4）。

RNA 的骨架和 DNA 的有些不同——核糖代替了脱氧核糖，并且 DNA 中 4 种碱基之一的 T 被替换成了略有不同的 U；其他三种碱基，A、C 和 G，则是相同的，RNA 中的所有 4 种碱基在 DNA 中都有互补的对应碱基——RNA 中的 A、U、C 和 G 分别对应于 DNA 中的 T、A、G 和 C。转录时，RNA 的单体成行排列，并以与 DNA 单体在复制中被选择的相同方式而被选择用于在 DNA 模板链上进行聚合。因此转录输出的是一条忠实反应细胞内部分遗传信息的单链多聚分子，只不过输出的字母——RNA 单体和原 DNA 单体有少许的不同。

同一段 DNA 分子可以重复的合成多条相同的 RNA 转录产物。因而，细胞的遗传信息储存形式 DNA 分子是固定的、不可变的，而 RNA 分子则可成批生产，并且用完之后可以丢弃（图 1-5）。我们知道，大多数这些转录产物的基本功能是作为遗传信息传递的中间体：它们以**信使 RNA**（messenger RNA，mRNA）的形式，依照 DNA 中储藏的遗传指令指导蛋白质的合成。

RNA 分子的特殊结构决定了它们拥有其他特定的化学功能。由于分子是单链的，并且骨架可以弯曲，所以 RNA 分子的一部分可以和另一部分微弱结合。当一条分子上的部分序列局部互补时，就会发生上述现象，如一段…GGGG…的序列就有可能和一段…CCCC…的序列结合。这种类型的内部结合力使得 RNA 分子折叠成一种特殊的空间构象（图 1-6）。反过来，RNA 分子的这种特殊结构能使它们选择性识别并结合到特定的分子上，并且在某些特殊的情况下可以催化其结合的分子发生化学改变。我们在本书后面将会看到，这种 RNA 分子的催化反应在细胞某些最原始和最基本的生理过程中发挥重要作用，并且有人设想在生命进化的

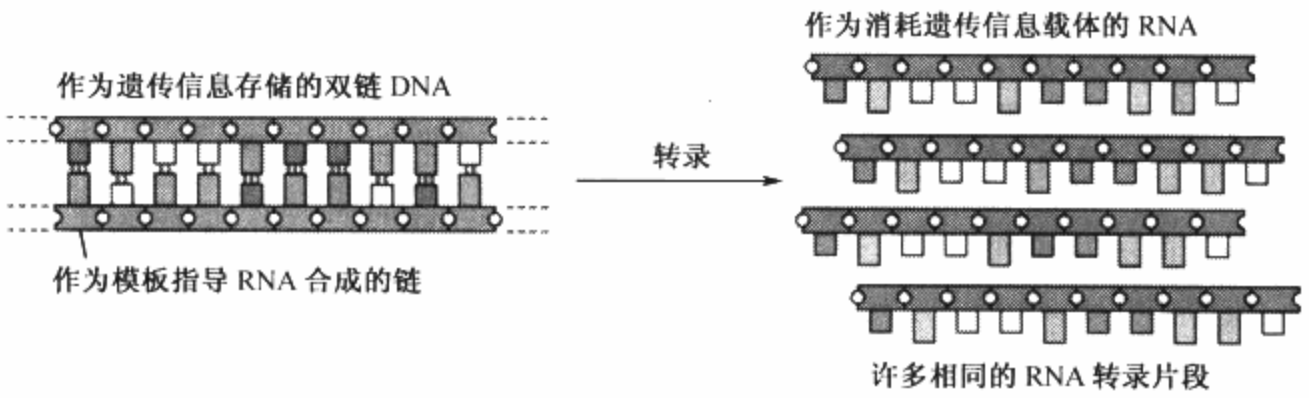


图 1-5　遗传信息如何在细胞内传递

每个细胞都有一系列的 DNA 分子——遗传信息的档案馆。特定的 DNA 片段指导许多相同 RNA 分子的合成，这些 RNA 分子作为档案馆中遗传信息的工作拷贝。通过转录一条长 DNA 分子的不同部分产生许多不同的 RNA 分子，从而使得细胞可以有选择地利用自身的遗传信息

初期这种 RNA 的催化反应更为广泛，并起着核心作用（讨论见第 6 章）。

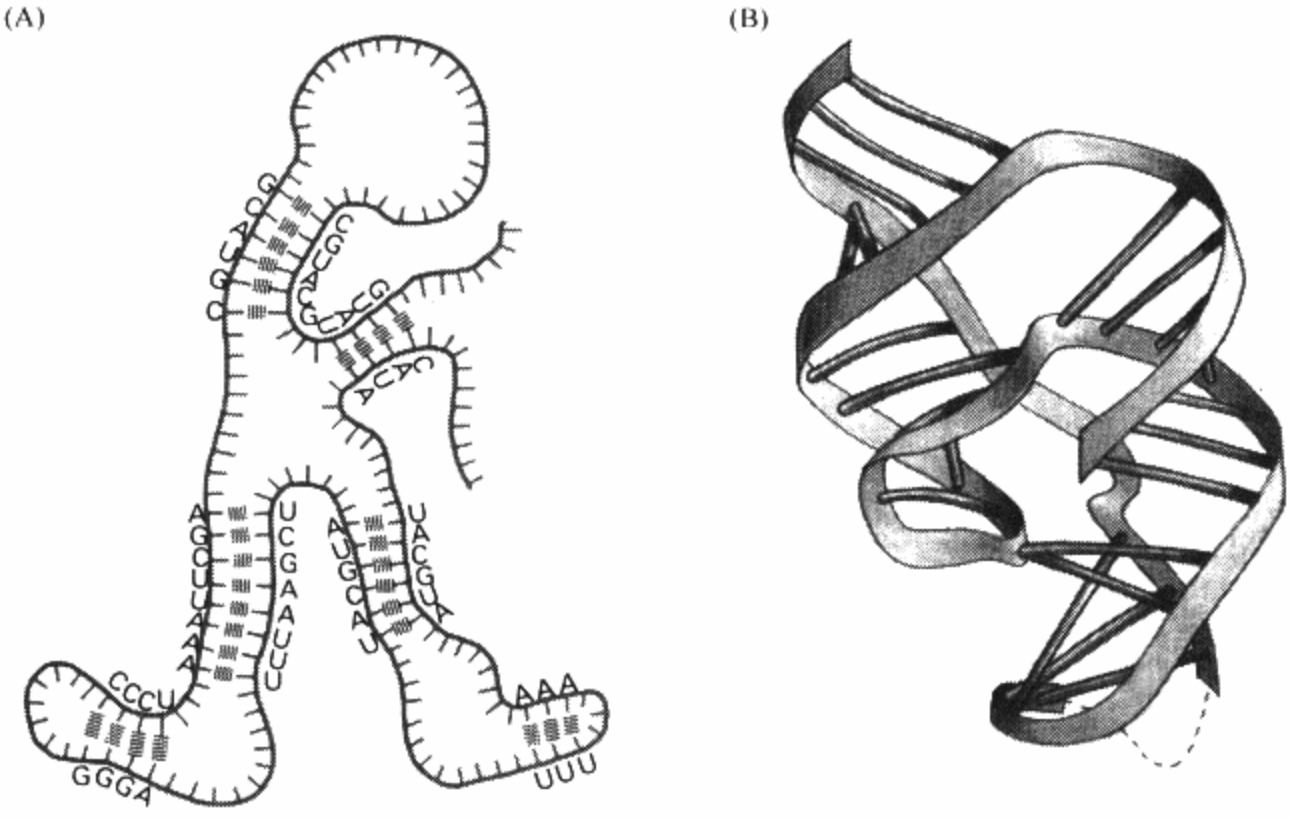


图 1-6　RNA 分子的结构

（A）同一个分子不同部分的碱基配对使得 RNA 分子有着特征性的外形。（B）一个丁型肝炎病毒的 RNA 分子的三维结构，具有催化 RNA 链断裂的功能。（（B），based on A. R. Ferré D'Amaré, K. Zhou, and J. A. Doudna, *Nature* 395: 567—574, 1998.）

所有细胞都将蛋白质用作催化剂

和 DNA、RNA 一样，**蛋白质**分子也是没有分支的聚合长链，由所有活细胞内相同基本组分中的单体聚合而成；并且蛋白质也是以线性序列的形式承载信息，就像人类使用字母一样。每个细胞内都有许多不同的蛋白质，除去水之后，它们占据了细胞质量的大部分。

蛋白质的单体——**氨基酸**（amino acid）与 DNA 和 RNA 的单体不同，它们共有 20 种，而不是 4 种。每种氨基酸都有一个相同的核心结构，并通过这一核心结构以标准方式与任何其他的氨基酸连接，连在核心上的侧基决定了氨基酸不同的化学特性。每个蛋白质分子或**多肽**（polypeptide）都由相互连接在一起的氨基酸序列形成一个精密的三维结构，其活性位点位于表面［图 1-7（A）］。因此这些氨基酸聚合物可以高度特异地结合其他分子，发挥**酶**

(enzyme) 的作用，催化有共价键的形成和破坏的反应。它们通过这种方式指导细胞内的绝大部分化学过程［图 1-7 (B)］。除此之外，蛋白质还有许多其他的功能——维持结构、产生运动、感知信号等。每种蛋白质由于其自身的特定氨基酸序列而行使特定的功能。最重要的是，蛋白质是可以把细胞的遗传信息赋之于行动的分子。

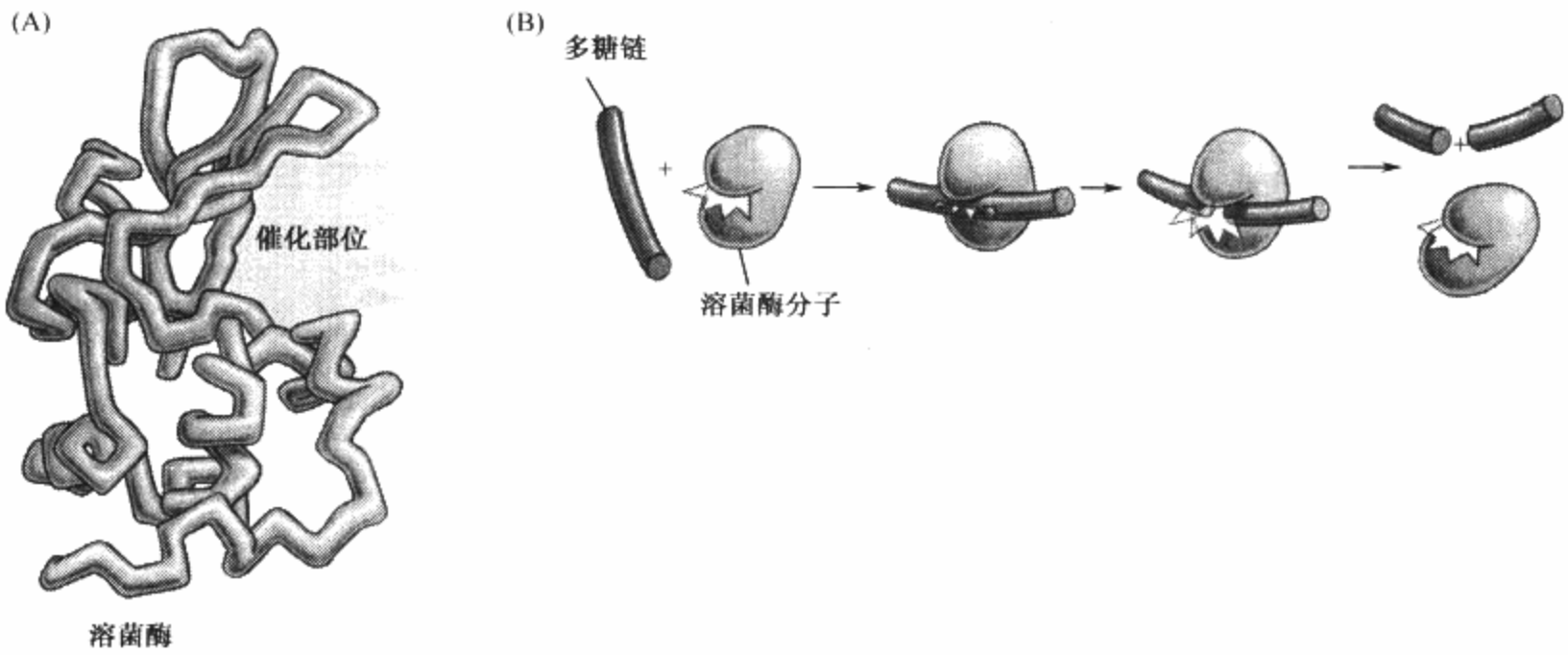


图 1-7　蛋白质分子怎样在化学反应中起催化作用

(A) 氨基酸序列决定蛋白质分子折叠形成一个特殊的形式。溶菌酶表面形成一个沟状催化位点。(B) 多糖分子（红色）是一个单糖分子的聚合链，结合到溶菌酶的催化位点，由于沟槽中氨基酸具有催化共价键断裂的作用而被断开

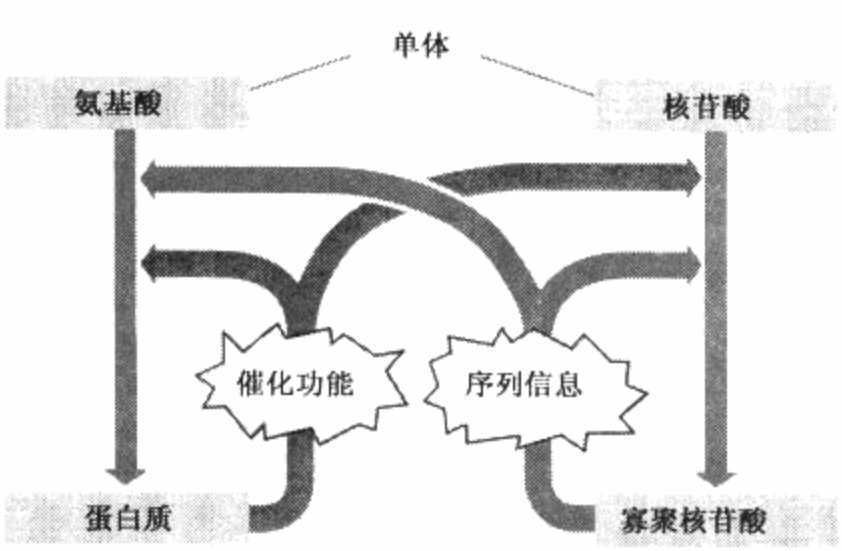


图 1-8　生命是一个自我催化的过程

寡聚核苷酸（核苷酸聚合物）和蛋白质（氨基酸聚合物）提供了序列信息和催化功能，通过一系列复杂的化学反应，可以合成更多的同种类型的寡聚核苷酸以及蛋白质

总而言之，多聚核苷酸决定蛋白质的氨基酸序列。反过来，蛋白质可以催化许多化学反应，包括 DNA 分子合成，遗传信息由 DNA 传递到 RNA 和蛋白质。这种反馈回路是机体自动催化、自我复制的基础（图 1-8）。

所有细胞都以相同的方式将 RNA 翻译成蛋白质

遗传信息从由 4 种字母组成的多聚核苷酸翻译成由 20 种字母组成的蛋白质是一个复杂的过程。在某些方面，这种翻译的规则简洁而合理，但另一方面，如果遇到这些规则在所有的生物体中（少量例外）都相同，那它们又是极为武断的。现在认为，那些武断的特征反应了生命起始初期的残酷意外事件——这些最早期有机体的可能特征被遗传下来，并被根深蒂固地植入细胞内而无法改变，除非细胞结构被破坏。

遗传信息以 3 个核苷酸一组的形式录入信使 RNA 的序列中：这种三核苷酸形式又叫密码子，决定（编码）了对应蛋白质中的一个氨基酸。由于存在 64 (4×4×4) 种可能的密码子，而仅有 20 种氨基酸，所以出现了几个密码子对应同一种氨基酸的情况。一类特殊的小 RNA 分子家族，**转移 RNA**，可以读出这些编码。每种类型的 tRNA 都是一端连接一个特定

的氨基酸，另一端为 3 个核苷酸组成的特殊序列（反密码子），它可以通过碱基配对特异性地识别 mRNA 上的密码子单元（图 1-9）。

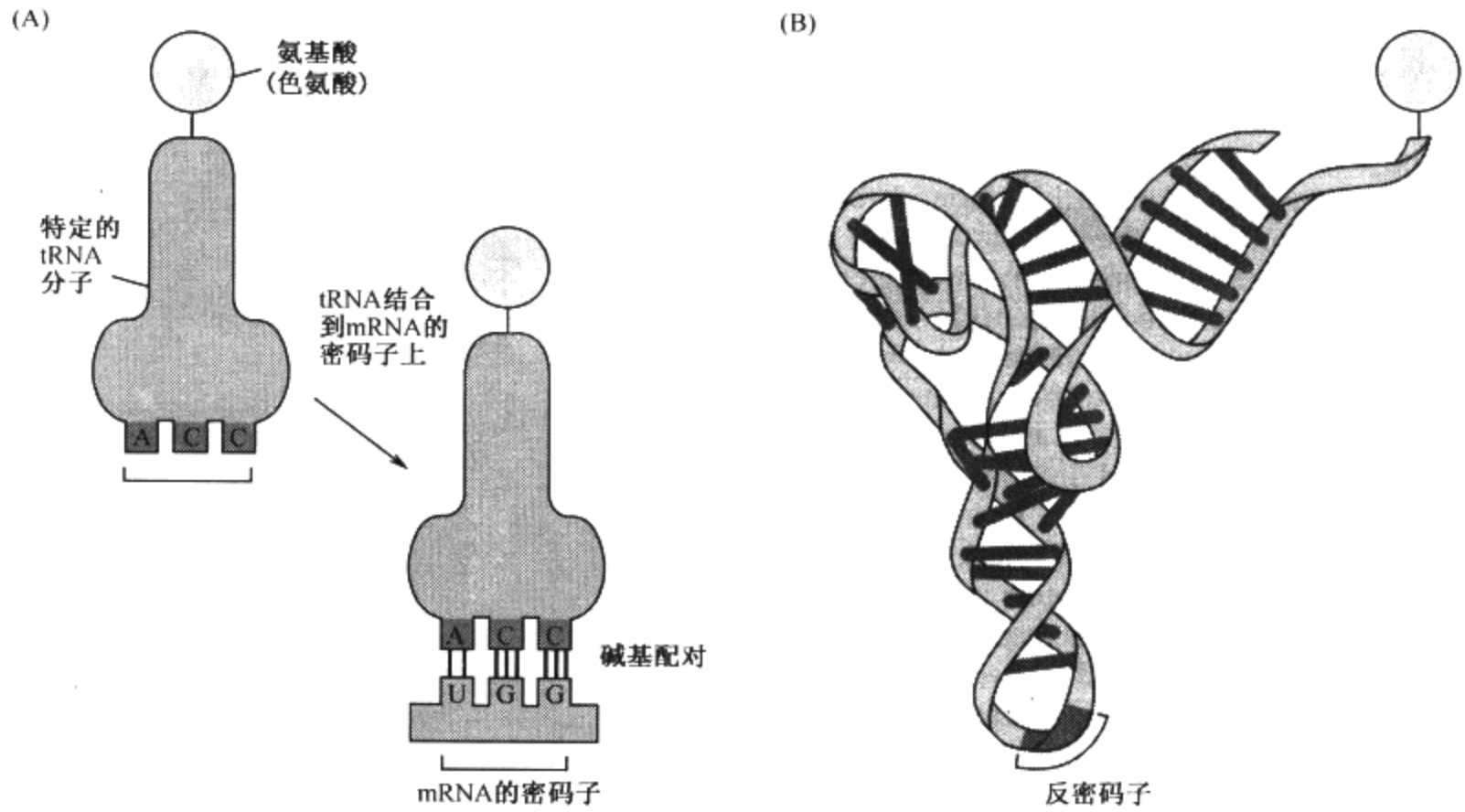


图 1-9　转移 RNA

（A）对应色氨酸的 tRNA 分子。tRNA 分子的一端带有色氨酸，另一端有三链核苷酸 CCA（反密码子），可以识别信使 RNA 分子上的色氨酸密码子。（B）色氨酸 tRNA 分子的三维结构。注意（A）中密码子和反密码子是反方向的，就像 DNA 双螺旋中的两条链一样（图 1-2），所以 tRNA 中的反密码子序列是从右往左读，而 mRNA 的密码子序列则是由左往右读

蛋白质合成时，一系列的 tRNA 分子通过反密码子与 mRNA 上的密码子进行碱基配对，从而把它携带的氨基酸连接到一起。当氨基酸连接到一起形成蛋白质链时，完成任务的 tRNA就脱落下来。这个复杂的过程是由一种庞大的多分子机器核糖体来完成的。核糖体由两条**核糖体 RNA（rRNA）**主链和 50 多种不同的蛋白质组成。这种历经了久远年代的分子紧紧抓住 mRNA 分子的末端并沿其滑动，其间捕获携带有氨基酸的 tRNA，把氨基酸串成一条新的蛋白质链（图 1-10）。

相应于一种蛋白质的遗传信息片段就是一个基因

通常，DNA 分子非常庞大，含有成千上万种蛋白质的信息。因此，DNA 序列的不同部分可以转录成不同的 mRNA 分子，每个分子编码一种蛋白质。**基因**就是对应于一种蛋白质的 DNA 序列（或者对应于催化 RNA 和结构 RNA 的 DNA 序列，因为它们的最终产物是 RNA 而不是蛋白质）。

细胞中单个基因的表达是受调控的：细胞可以根据需要自动调控转录和翻译的速率，而不是所有的时间都全力合成蛋白质。调控 DNA 分散在编码基因之间，这些 DNA 非编码区可以结合特定的蛋白质，从而调控局部基因的转录速率（图 1-11）。另外，还存在一些其他的非编码 DNA 序列。例如，有些 DNA 的作用就像标点符号，可以决定某种蛋白质在哪里起始，

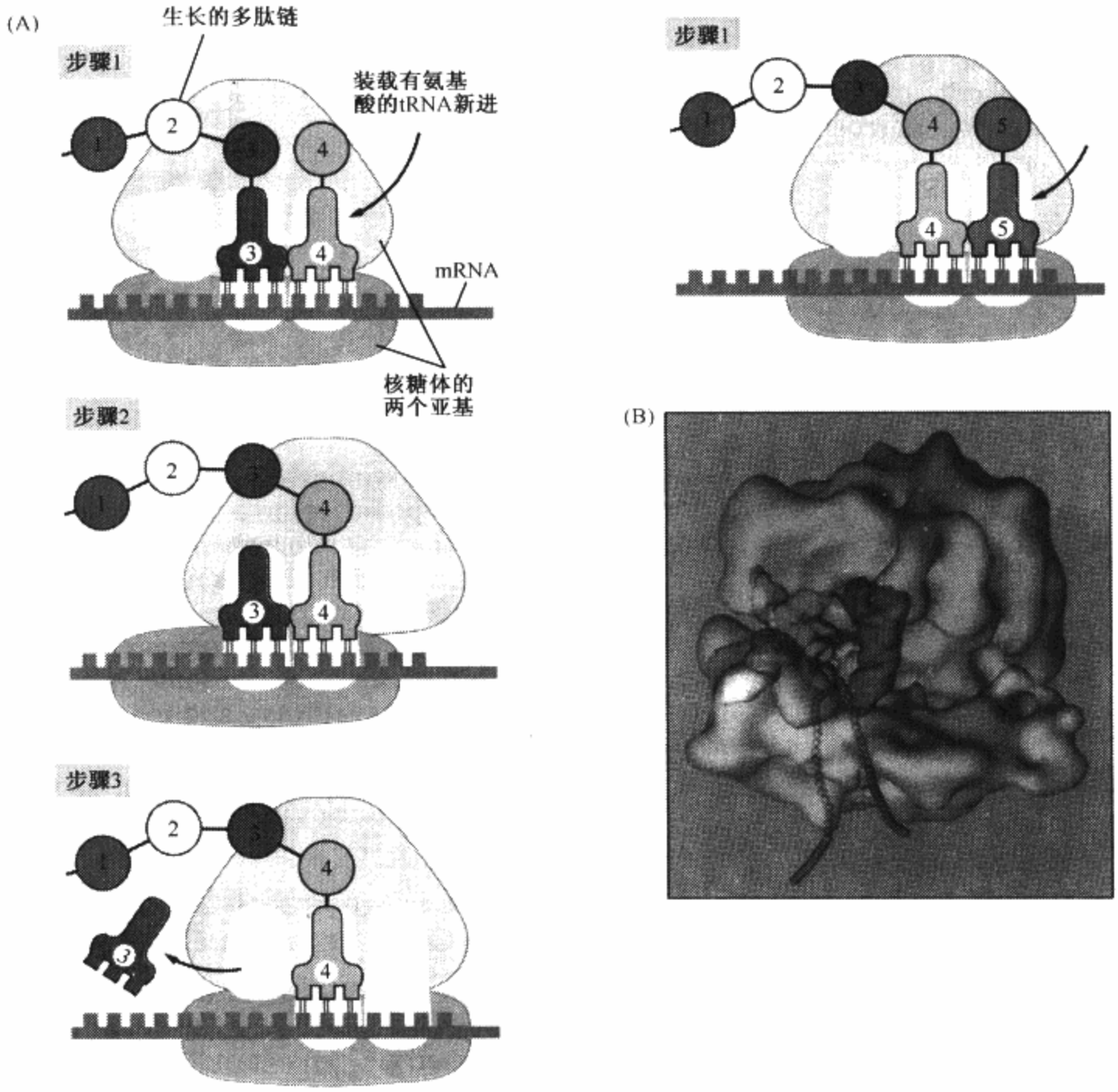


图 1-10　工作中的核糖体

（A）此图显示了核糖体怎样沿 mRNA 分子移动，怎样捕获与 mRNA 的密码子对应的 tRNA，并把其携带的氨基酸组成蛋白质序列。mRNA 决定了氨基酸序列。（B）细菌核糖体的三维结构（淡绿色和蓝色），沿着 mRNA 分子（橙色小珠）移动，有 3 个在捕获和释放不同步骤中的 tRNA 分子（黄色、绿色和粉红色）。核糖体是一个含有 50 多种不同蛋白质以及 RNA 分子的庞大装配车间［（B）经 Joachim Frank，Yanhong Li 和 Rajendra Agarwal 许可］

到哪里结束。这些调控 DNA 和其他功能的非编码 DNA 在各个组织器官中的数量和分布都不相同，但它们的基本功能都大体相似。总之，细胞的**基因组**——就是所有的 DNA 序列中所含有的全部遗传信息——不仅决定细胞内的蛋白质性质，还控制它们表达的时间和方位。

生命需要自由能

一个活细胞远不是一个处于化学平衡的系统：内部存在着大量的自由能，即如果细胞朝着化学平衡的方向进行衰退或凋亡，它会向环境中释放大量的热量。如果细胞要依照自己的模式复制出一个新的细胞，它必须像吸取原材料一样从环境中吸取自由能，以驱动必要的合成反应。这种自由能的消耗对生命来说是最基本的，如果它停止了，那么生命就会死亡。我们知道遗传信息也是生命的基本元素，那么它们之间有什么联系呢？

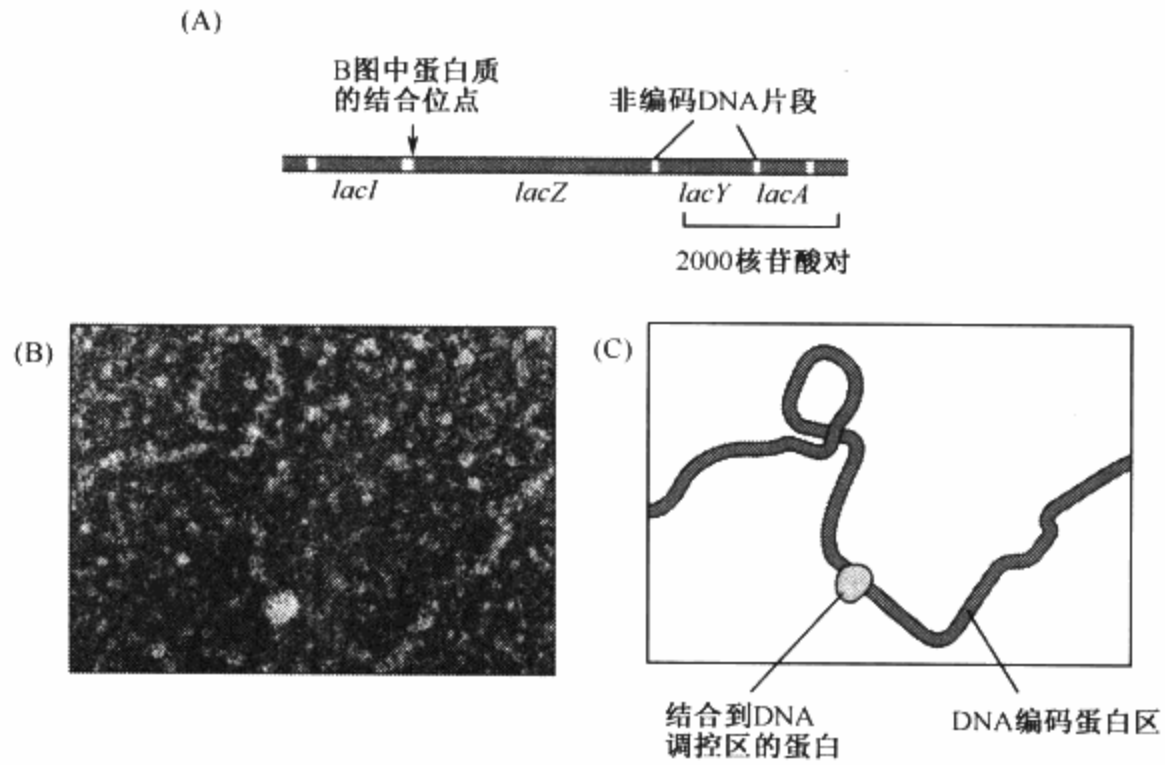


图 1-11　(A) 大肠杆菌的一小部分基因组示意图，含有编码 4 种不同蛋白质的基因 (*lacI*、*lacZ*、*lacY* 和 *lacA*) 编码蛋白质的 DNA 片段（红色）之间有调控区域以及其他非编码 DNA（黄色）。(B) 这一区域的电镜图片，有一个蛋白质分子（由 *lacI* 基因编码）结合于调控区；此蛋白质可以调控 *lacZ*、*lacY* 以及 *lacA* 基因的表达。(C) (B) 中结构的草图［(B) 经 Jack Griffith 许可］

答案是肯定的，遗传信息的继承需要自由能，并且它们之间有一个精确的量化关系。要了解一些信息，即在两种有相同可能性的选项间选择是或否，就要消耗一定量的自由能（以焦耳计算），量的大小取决于温度。证明这种抽象的统计热力学普遍原理极为困难，并且还依赖于“自由能”的精确定义（讨论见第 2 章）。然而，在 DNA 合成的背景下，这一基本概念还是不难理解的。

要产生一条和已有 DNA 分子相同的新 DNA 序列，核苷酸单体必须正确地排列在现有的 DNA 模板上。在序列的每个点上，适当核苷酸的选择依赖于以下事实，正确配对的核苷酸和模板的结合力大于错误配对的核苷酸与模板的结合力。结合力越大，错误结合的概率就越小。一个高保真的匹配，无论是通过我们刚刚提过的简单机制直接形成的，还是通过一系列复杂化学反应形成的，在正确核苷酸插入时都伴随着大量自由能的释放。这一切除非是在起始的分子带有高储量的自由能时才不会发生。最后，新的核苷酸结合在一起形成新的 DNA 链，又需要新的自由能把新链从旧链上分离下来以进行下一轮复制。

因此，细胞为了忠实的复制自身遗传信息，需要从所处的环境中吸取自由能。同样这条规则也适用于细胞合成其他分子。例如，在 RNA 和蛋白质合成时，现有的遗传信息通过一系列的分子匹配指引新分子的生成，此时需要自由能驱动多种化学分子的反应，从原材料中构造出单体并把它们正确地连接到一起。

所有的细胞都是有着相同分子建造材料的生化工厂

由于细胞要合成 DNA、RNA 和蛋白质 ，而每一种大分子又是由一系列的亚单元组成，所以所有的细胞必须处理一批相似的小分子，包括单糖、核苷酸、氨基酸以及其他合成时普遍需要的物质。例如，所有的细胞都需要磷酸核苷酸 ATP（三磷酸腺苷）作为 DNA 和 RNA

合成时的建筑块；再如，所有的细胞都需要制造和消耗这种携带自由能和磷酸基团的载体分子以驱动许多其他的化学反应。

虽然各种细胞都是相似型号的生化大工厂，但它们对于一些小分子的处理细节则有很大的差异，信号大分子的特性也不是完全相同的。像植物等一些机体仅仅需要最简单的营养素和太阳光的能量就可以制造出自身的有机小分子，而像动物等另一些机体则必须摄入现有的生物体并从中吸取它们现成的有机分子。我们将在下面讨论这一点。

细胞外被覆一层细胞膜，营养物质和废弃物必须通过细胞膜进出细胞

除了上述的特点以外，细胞至少还有一个共同点：每个细胞都被覆一层膜——**细胞膜**。这层膜作为一个选择性的屏障可以聚集从环境中收集的营养物质，保证自身合成的产物为己所用，同时往外分泌出废弃物。如果没有细胞膜，那么细胞就不可能作为一个协同的化学系统而维持其完整性。

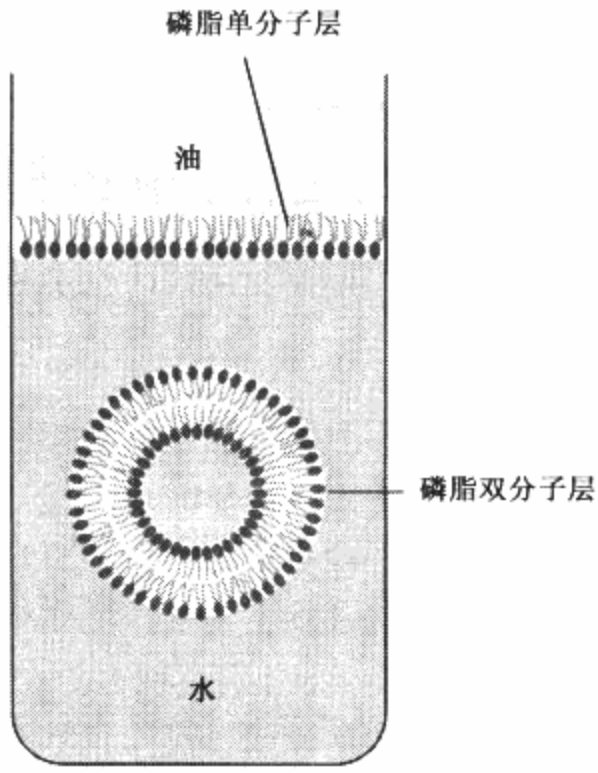


图 1-12　两亲性磷脂分子形成膜结构

分子都含有一个亲水头部（磷酸基团）和疏水尾部（碳氢化合物）。在水和油的分界线上，它们排成一排，头部基团向水，尾部向油。浸入水下时，它们聚合在一块形成内部水质的双层膜结构

细胞膜是由一系列具有双亲性理化性质的分子组成，即一部分是疏水的（水不可溶的），另一部分是亲水的（水可溶的）。当这些分子放置在水中时，它们可以自发的聚集，疏水部分尽可能聚在一起，隐藏在内部，而亲水集团则暴露在外。像磷脂分子之类的适当形状的亲水脂分子占据了细胞膜的大部分，它们在水中自发形成紧贴的双层囊泡结构（图 1-12）。可以做个试验来观察这种现象，在一个试管中混合磷脂和水，在适当的条件下，小囊泡的水层结构可以明显地和外层培养基相区分。

虽然所有细胞中主要膜分子的化学结构不同，但其疏水尾都是碳氢聚合物（$—CH_2—CH_2—CH_2—$），并且它们自发形成双层囊泡只是许多普遍规律中的一个例子：细胞合成的化学分子的化学特性促使它们自组装形成细胞所需要的结构。

细胞的这层边界并不是完全不可透过的。当细胞在生长和分裂时，它必须要透过细胞膜从外界吸取营养物质，并排出自身废弃物。因此细胞膜上都结合有特定的蛋白质，用于将特殊的分子从一端运送到另一端（图 1-13）。一些膜输送蛋白质和细胞内催化基本小分子反应的蛋白质一样，在进化过程中是非常保守的，所以人们可以通过对比两个物种（即使是进化上差别很远的）来认识这个蛋白质家族的相似性。

膜上的转运蛋白质主要决定进入细胞的分子类型，而细胞内的催化蛋白质决定了这些分子所要参与的反应。因此，通过指定细胞内所要合成的蛋白质，DNA 分子上的遗传信息不仅指挥着细胞内所有的化学反应，还控制着细胞的形态和行为，因为它们也主要由细胞的蛋白质构建和控制。

细胞的生存仅仅需要不到 500 个基因

生命信息传递的基本规律非常简单，那么实际的细胞内有多么复杂呢？尤其是细胞的最

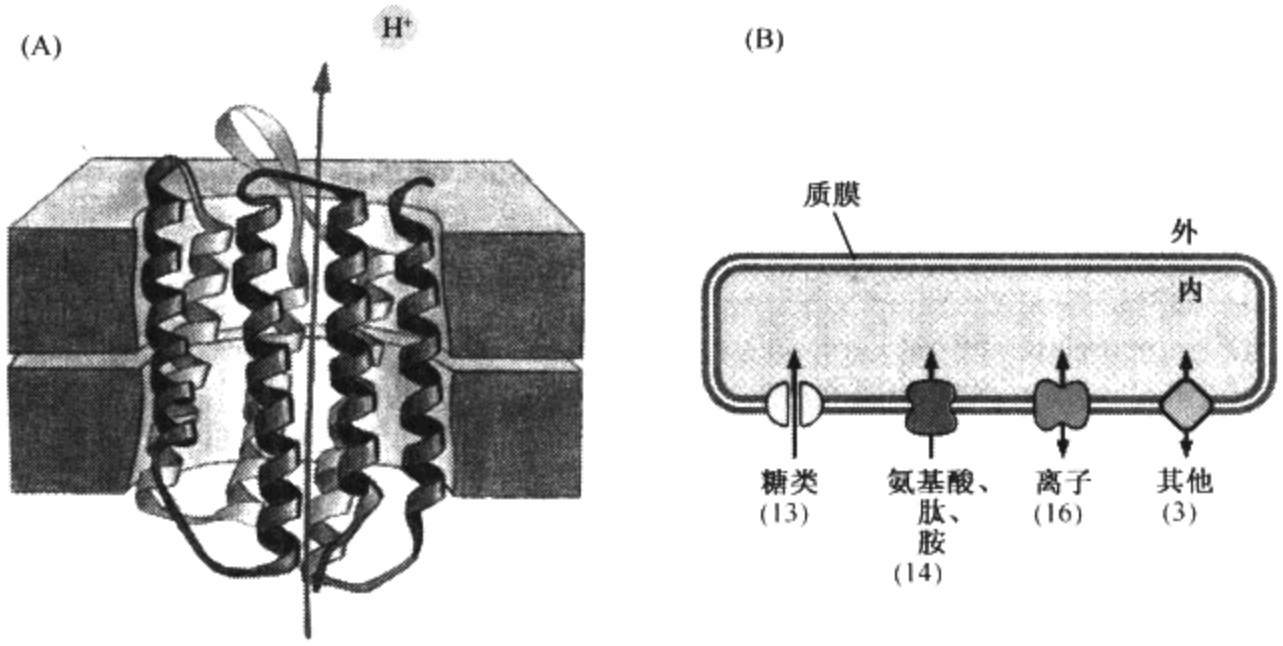


图 1-13　膜转运蛋白质

（A）古细菌 *Halobacterium halobium* 紫膜质的分子结构。这种转运蛋白质可以利用吸收的光能把离子（H^+）泵出细胞。多肽链呈螺旋状穿过细胞膜；在一些区域缠绕呈螺旋状，螺旋片段排列成离子转运通道的管壁。（B）细菌 *Thermotoga maritima* 细胞膜上转运蛋白质示意图。数字代表每种不同类型转运蛋白质的数量。大部分同种类型蛋白质在进化上彼此类似

低需求有哪些呢？关于这些，我们可以通过研究一个基因组最小的物种——支原体，而得到一些粗略的结论（图 1-14）。这种细菌寄居在哺乳动物体内，所处的环境提供给了它许多现成的小分子。但它还必须自己生成所有基本遗传所需的大分子——DNA、RNA 以及蛋白质。支原体仅有 477 个基因，基因组由 580 070 个碱基对组成，代表 145 018 个字节的信息——相当于本书中一章所包含的字数。它的细胞生物学可能很复杂，但并非不可能生存。

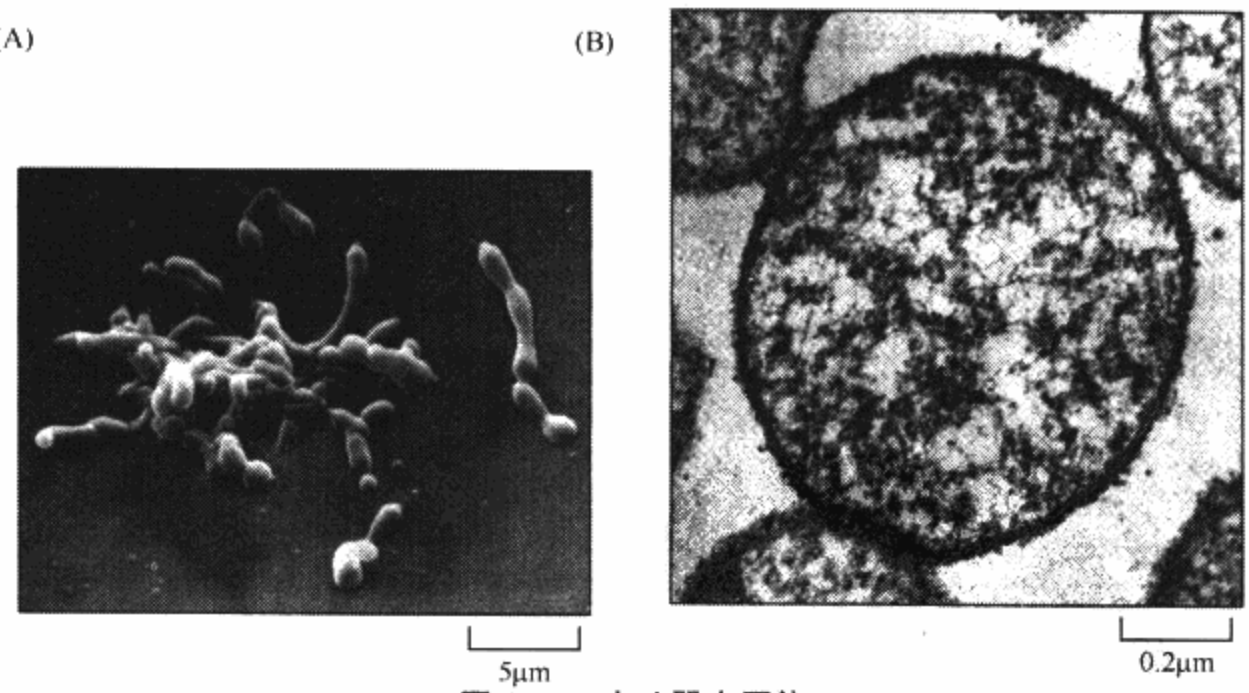


图 1-14　生殖器支原体

（A）扫描电镜显示这种小细菌的不规则外形，它缺乏坚硬的细胞壁。（B）支原体细胞横切面（发射电镜照片）。在生殖器支原体的 477 个基因中，37 个编码转运 RNA、核糖体 RNA 和非信使 RNA。297 个编码基因的功能已知或预测到：其中 153 个在 DNA 复制、转录、翻译以及 DNA、RNA 和蛋白质相关的步骤中起作用；29 个与细胞膜表面结构有关；71 个在能量转换以及小分子的合成和降解中起作用；11 个在细胞分裂调控以及其他步骤中起作用（A，from S. Razin, M. Banai, H. Gamliel, A. Pollack, W. Bredt, and I. Kahane, *Infect Immun.* 30: 538-546, 1980; B, Courtesy of Roger Cole, in Medical Microbiology, 4th edn., [S. Baron ed.]. Galveston: University of Texas Medical Branch, 1996.）

目前所知活细胞内需要的基因数目至少为 200～300 个。.下一节将会提到，如果将进化树

上的分支进行比较，我们会发现有 200 多个基因在所有物种中是相同的。

小结

有机体通过把遗传信息传递给子代来复制自身。单个细胞是最小的自我复制单元，并且在所有物种中都作为遗传信息的传递载体。地球上每个细胞都把自身的遗传信息以相同的化学形式存储，即双链 DNA。信息复制时，DNA 双链打开，每条链都可作为模板链合成与自身完全互补的新 DNA 链。转录时，DNA 采取同样的依赖于模板的聚合方式把部分信息转化成相似的聚合物——RNA。RNA 接着又通过更为复杂的翻译机械装置指导蛋白质的合成，其中包括一个由 RNA 和蛋白质组成的多分子机器——核糖体。蛋白质是细胞内几乎所有化学反应的催化剂；它还有其他的功能，包括可以选择性的输入、输出小分子进出细胞膜。蛋白质的特有功能取决于它的氨基酸序列，氨基酸序列又取决于编码蛋白质的 DNA 序列——基因。依照上述方式，细胞的基因组决定体内的化学反应，而由于细胞必须合成 DNA、RNA 和蛋白质，所以每个细胞都含有相似的基本生化过程。已知的最简单细胞内仅有不到 500 个基因。

基因组的多样性及生命树

活体是具体可见的，而不像其他的化学形式那样给人以无尽的猜想。它们占据了海洋，覆盖了陆地，浸润了地球的外壳，繁荣了地球的表面。富集氧的大气层、煤炭以及石油，多层的铁矿石、白垩、石灰岩以及大理石形成的悬崖，所有的这些，都直接或间接地来源于古生物的活动。

生物体不受环境限制，生活在温和的陆地、水底以及阳光充足的地方。它们可以生存在海洋的最深处、火热的火山口、南极洲冰封地面下的水域里以及地球表面下数千米的深处。这些生活在极端环境下的生物都是不常见的，不仅因为它们所处的环境难以到达，而且它们大多数是肉眼看不到的。即使在平常的地方，大多数生物体也是我们不使用特殊仪器所无法看到的：它们往往不引人注意的运动着，除非它们导致一场疾病或者腐蚀掉我们家房屋的木材。但是微生物却占据了我们星球所有生物体总质量的大部分。直到最近，通过新的分子分析方法甚至 DNA 序列分析，我们才开始得到一副地球生命的画卷，显示出并不是我们以前所歪曲的只有大型动物生存在陆地上的景象。

这一节里我们将讨论生命的多样性以及它们之间的联系。由于所有物种的遗传信息都是以 DNA 序列的形式存在的，并且目前已经通过生化技术破译了部分物种的 DNA 全序列，所以我们有可能参照它们的序列进行分析、存档或者比较任意物种。通过对比我们可以评估某一物种在整个生命家族树中所处的位置——这就是“生命树”。在我们描述这项方法的意义之前，我们必须先了解不同环境下的细胞是怎样获得它们生存繁殖所必需的物质的，另外还需了解一些种类的机体必须依赖另一种类的物种提供基本化学物质的方式。

细胞可由多种自由能源提供能量

生物体可以通过多种方式获得自由能。像动物一样，真菌以及生活在人类消化道中的细菌通过摄取其他活的机体或它们产生的化学有机物获得自由能；这些物种被称为有机营养生物（来源于希腊语 trophe，意为“食物”）。另外一些从非生物界吸取能量。这可分为两类：一类可以利用太阳能，另一类可以由环境中无机化学物质中的高能量系统中获得能量（远离

化学平衡的化学系统）。前者我们叫做光养性生物（以太阳光为食），后者叫做石养性生物（以石头为食）。有机营养生物离开这些能量转换器是无法生存的，后者占据了地球生命物质的大部分。

光养性生物包括多种细菌、藻类以及植物，是我们人类，甚至我们通常所看到的生物，所依赖的。光养生物改变了我们环境的化学构成，地球大气层的氧气是它们生物合成活动的副产物。

石养性生物在我们的世界里没有明显的特征，因为它们都是微观的，并且大多数生活在我们人类到达不了的地区——海底深处、地下数千米以及其他许多极端环境。但它们是生命世界的重要组成部分，甚至曾经在地球生命发展的过程中起着重要作用。

一些石养性生物从需氧反应中得到能量，它们利用环境中的氧分子。由于大气中的O_2是活体的最终产物，所以在某种意义上说，这些需氧的无机营养菌消耗的是以前生物的产物。然而，其他的厌氧性无机营养菌可以生活在极少或者没有氧分子存在的环境，这种环境和地球形成早期没有氧气积累的环境差不多。

这些环境中最不可思议的是在太平洋和大西洋洋底地层上的热气口，地球内部物质从那里逐步涌出并扩展为新的海底陆面（图 1-15）。渗入的海水被热浪卷出，形成一个海底喷泉，水中携带有地球内部的化学物质。比较典型的有H_2S、H_2、CO、Mn^{2+}、Fe^{2+}、Ni^{2+}、CH_2、NH_4^+以及含磷化合物。有大量的细菌生活在这个排气口周围，品味这种涩味的饮食，从化学反应中捕获自由能。其他的物种，放线菌、蚌类、巨蟾的卵，依次排列在细菌之外，形成

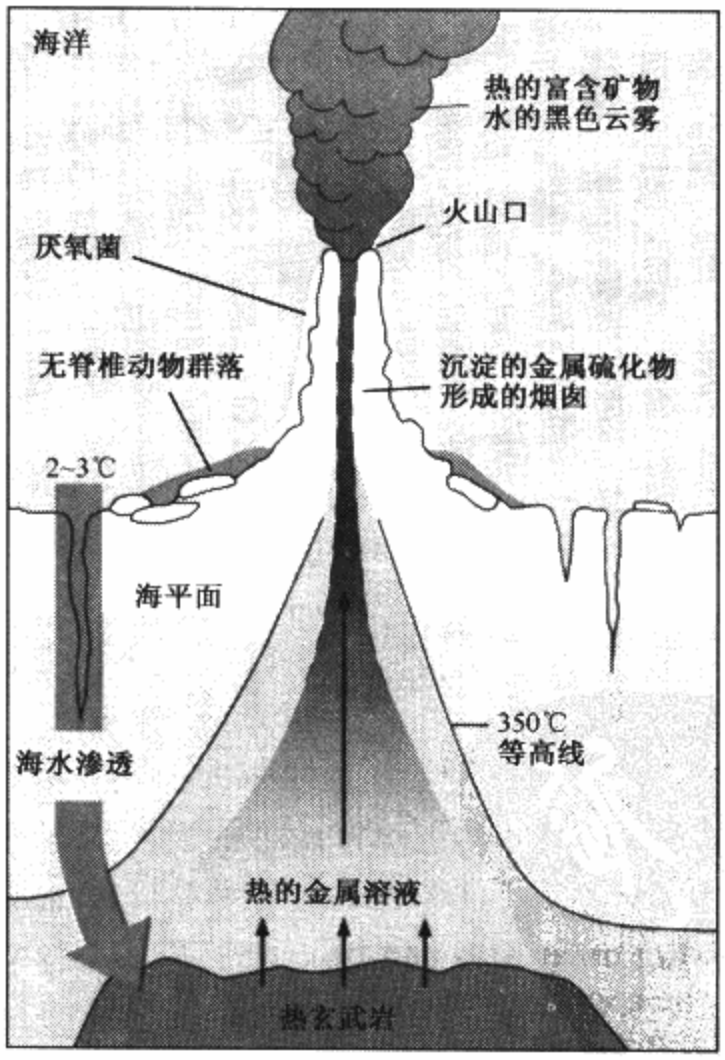


图 1-15　海底地层上热气口的地理面貌

渗入的海水被热浪卷出，形成一个海底喷泉，水中携带有地球内部的化学物质。此处有一个温度的梯度分布，由喷气口中心地带的 350℃以上，降低到海水接触面的 2～3℃。矿物质遇冷后沉淀为烟囱状。不同种类的生物生活在不同的温度下，在烟囱周围形成不同的领域。一般的烟囱状沉积物可达几米高，流速为 1～2m/s

一个类似于植物和动物的完整生态系统，但它们是以地球化学能而不是以光为能源（图 1-16）。

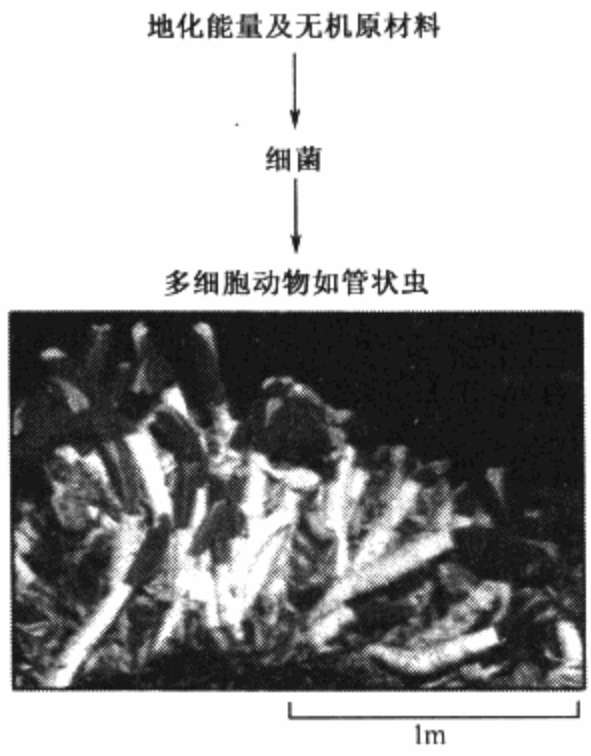


图 1-16　生活在热气口的生物

在气口周围温度可达 150℃以上，生存着多种石养的真细菌和原细菌，它们可以化学能量为食物。稍远的地方，温度稍微降低，生活着多种无脊椎动物，它们以这些微生物为食。最引人注目的是庞大（2m）的管状虫类，它们不以这些石养细胞为食，而是靠与它们的共生来生活：这种虫类的某种器官聚集了惊人数量的共生硫氧菌。这些细菌可以吸收地球化学能并把能量供给没有嘴、肠道以及肛门的宿主。管状虫依赖于收获地球化学能的细菌就像植物依赖于叶绿体捕获太阳能一样，我们会在以后的章节中讨论。一般认为，管状虫是由一种非常常见的动物演化而来的，现已成为热气口的第二大类生物（经 Dudley Foster，Woods Hole Oceanographic Institution 许可）

一些细胞可以为其他细胞固定氮及二氧化碳

细胞除了需要自由能之外，还需要其他的物质。DNA、RNA 和蛋白质仅由 6 种元素组成：氢、碳、氮、氧、硫和磷。这些元素广泛存在于地球上的非生物环境如岩石、水以及大气中，但是其存在形式并不容易直接形成生物活性分子。尤其是大气中的 N_2 和 CO_2 是惰性的，这些无机分子需要反应成有机化合物，以便为进一步的生物合成做准备，即将氮和二氧化碳固定，从而将 N 和 C 两种元素引入机体内，这一过程需要极大的自由能。许多物种的细胞缺少这种固定作用的生化机器，因此必须依赖于其他物种。动物依赖于植物提供有机碳和氮化合物。尽管植物可以固定空气中的二氧化碳，但它们没有固定氮气的能力，也必须依赖于固氮细菌为它们提供氮化合物。例如，豆科的植物根瘤中就存在着共生固氮菌。

因此，细胞的生化特性各种各样。不足为奇的是，在需要和功能互补的细胞之间有着紧密的关联。我们会在后续章节了解到，当这种关系发展到当其中一方完全丧失了某一特性的时候，互补的细胞就会形成单个复合细胞。

原核细胞间的生化多样性是最为广泛的

人们早已通过显微镜把细胞分为两大类：**真核细胞**（eucaryote）和**原核细胞**（procaryote）。真核生物的 DNA 存在于一个有膜边界的胞内隔室中，这一隔室被称为细胞核（来源于希腊语，意为“真正有核的”，*eu* 即“正确地”或“真的”，*karyon* 为“核的”或“仁的”）。原核细胞没有明显的核包裹 DNA。植物、真菌以及动物属于真核生物，细菌属于原核生物。

大多数原核细胞外表既小又简单，多以单体的形式而不是多细胞有机体的形式存在。它们的典型特征是球形或者杆状，线性长度为几微米（图 1-17）。它们通常有一件粗糙的保护性外套，叫做细胞壁，其内为细胞膜包裹的细胞质，细胞质由 DNA、RNA、蛋白质以及许多生命必需的小分子组成。电镜下可以看到，细胞内部没有可见的有规则的内部组织，内部构造是可变的（图 1-18）。

原核细胞占据一个庞大的生态位，令人惊奇的是它们的生化特性有着巨大的不同，这种不同远远大于真核细胞间的差异。有机营养性物种能以各种有机分子为食，包括糖、氨基酸、碳氢化合物以及沼气；光养性物种（图 1-19）能以各种途径捕获光能，其中一些还能以副产物的形式释放氧气；石养性物种能以简单的无机营养素为食，并从 CO_2 中获取碳源，依靠

图 1-17　一些细菌的外形和大小

虽然大部分细菌像图示的那么小，但也有一些庞然大物。一个非常极端的例子（没有给出）是雪茄状的细菌 *Epulopiscium fishelsoni*，它们生活在刺尾鱼的肠道里，大小可达 600μm 以上

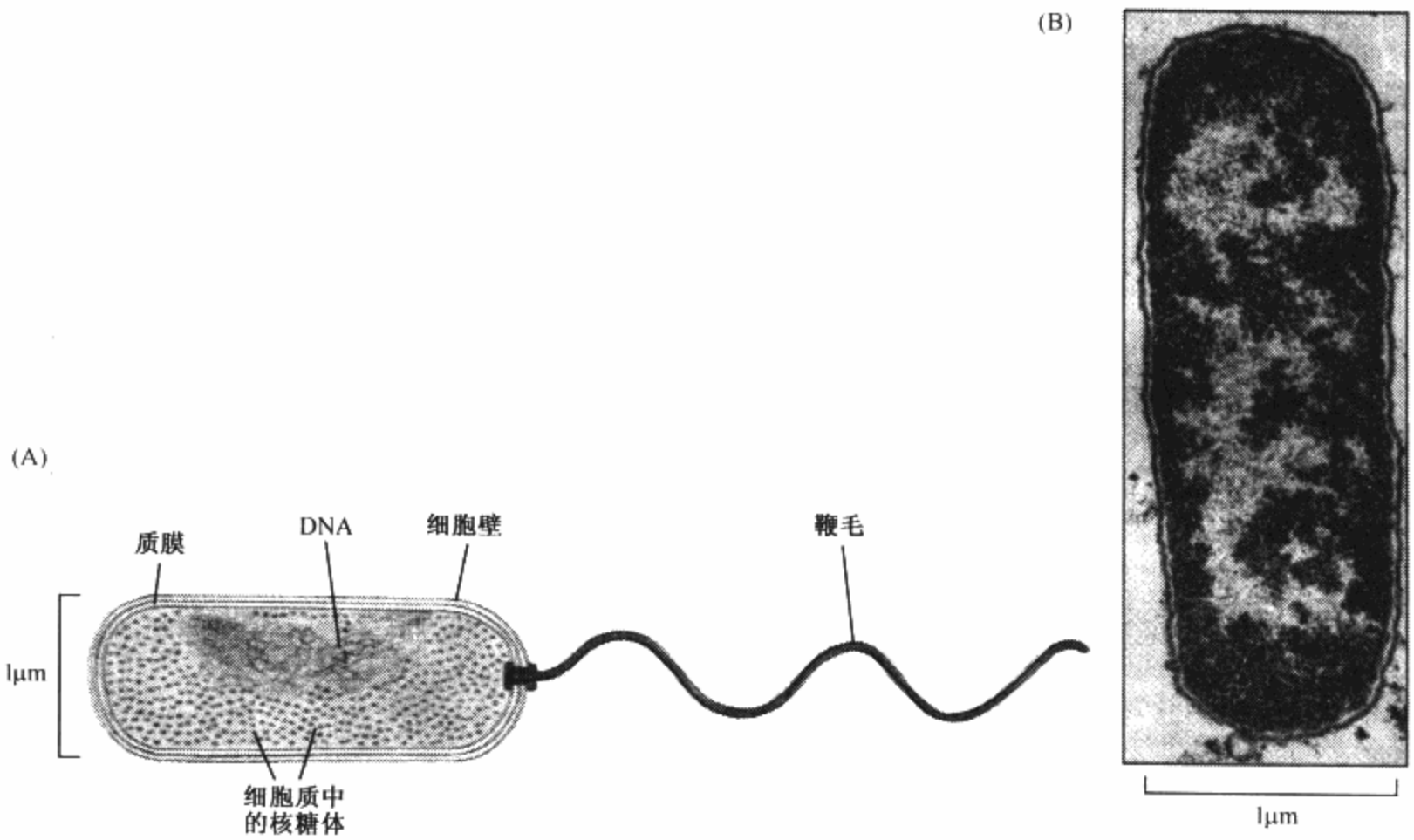


图 1-18　细菌的结构

（A）霍乱弧菌，显示出其简单的内部结构。和许多种属一样，霍乱弧菌一端有着螺旋状的附属物（鞭毛）它可以像推进器一样旋转来驱动细胞前行。（B）大肠杆菌（*E. coli*）的纵向电镜照片。它和霍乱弧菌相似但缺少鞭毛。细胞的 DNA 压缩在较亮的区域［（B）经 E. Kellenberger 许可］

H_2S 满足它们的能量需求（图 1-20）。另外还有 H_2、Fe^{2+}、硫元素以及环境中众多的其他化学物质。

事实上，世界上大多数微生物物种都没有被研究过。传统的细菌学方法可以让我们在实验室分离、培养一些菌种，从而了解它们的特性。但一些生活在天然环境下的细菌的 DNA 序列分析，如生活在土壤、海水甚至人类口腔中的细菌，使我们认识到大多数的细菌无法通过实验室技术进行培养。据估计，至少有 99%的原核生物未被定性。

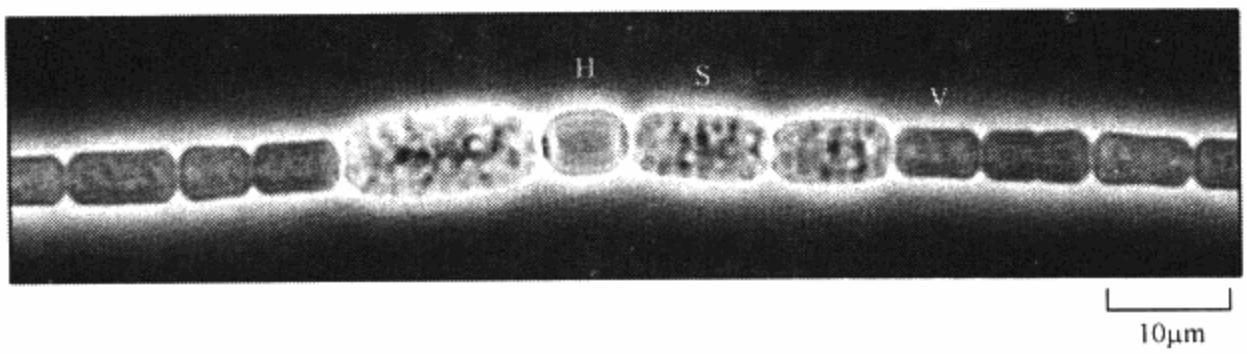


图 1-19　光养细菌 *Anabaena cylindrical* 光学显微镜下的照片

该物种细胞形成一个长的多细胞丝。大部分的细胞（标志 V）进行光合作用，其他的一些有着特殊的固氮作用（标志 H）或形成防御性的孢子（标志 S）（经 Dave G. Adams 许可）

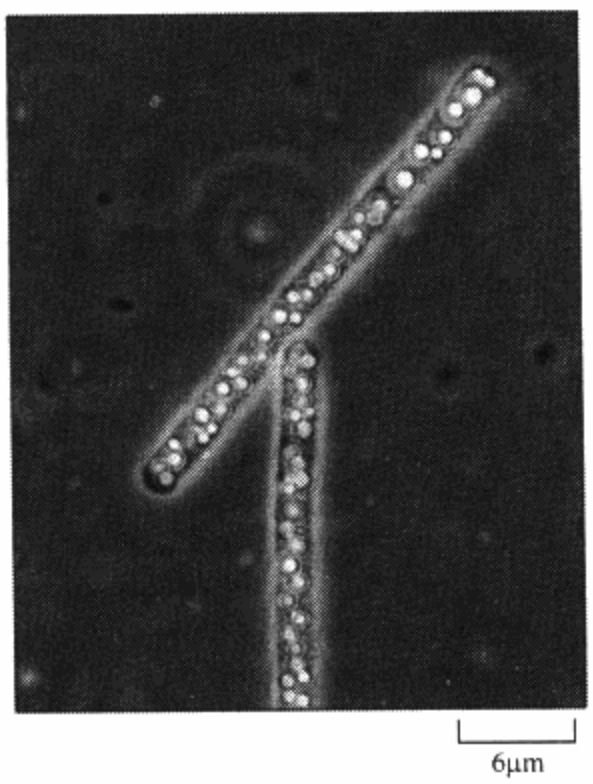


图 1-20　石养性细菌

Beggiatoa 生活在亚硫环境中，即使是在黑暗中亦可以从氧化 H_2S 中获取能量以及固定碳。注意细胞中硫磺的黄色沉积（经 Ralph W. Wolfe 许可）

生命树有三个基本分支：细菌、古细菌和真核生物

传统上，根据其外在表型对生物进行分类，我们可以看到，鱼类和我们一样有眼睛、颌骨、骨骼、脑等器官，但是蚯蚓却没有；蔷薇外表和果树相似，但和小草相差甚远。我们可以很容易地把这种相近家族的相似性解释为它们来源于一个共同的祖先，并且可以从化石中找出其祖先遗留的痕迹。依照这种方式，我们可以绘制出一个生物的家族树，可以清楚地表明各个遗传分支以及一种物种演化过程中区分于同一祖先其他物种的分支点。

然而，当物种之间差异太大时，这种方法就无法发挥作用了。我们怎么来确定一种真菌是否是一种植物或者动物的近亲呢？对于原核生物来说，这种方法也存在巨大缺陷：两种微生物有可能都是杆状的或球形的。因此微生物学家们寻找出了一种根据微生物生化性质和营养需求进行分类的方法。但这种方法也有自身的缺点。由于微生物有各种各样的生化特征，所以很难确定哪种区别真正反应它们在进化史上的不同。

基因组分析改变了这种困境，为我们分析物种的进化关系提供了一种简单的、更为直接也更有力的途径。物种的全长 DNA 序列可以为此物种的各个细节提供几乎完美的精确信息。并且这种信息我们曾经讨论过，是以数字的形式，即一串字符表现的，可以直接输入计算机并与其他物种的遗传信息进行比较。DNA 可以发生随机突变，并在长期的进化过程中积累（我们即将了解到），两物种间 DNA 序列的不同可以让我们直接、客观而定量地了解它们之间的进化距离。

人们利用这种方法发现，传统模式所划分一部分“细菌”在它们的进化点上和任何的真核生物以及原核生物都有着极大的分歧。表明原核生物由两种不同群体组成，它们在地球早期真核生物产生前就已产生分化形成两种不同的群体。这两种原核生物分别叫做**细菌**（或真细菌）和**古细菌**（或原细菌）。所以世界上的生物可以分为三种：真细菌、古细菌和真核生物（图 1-21）。

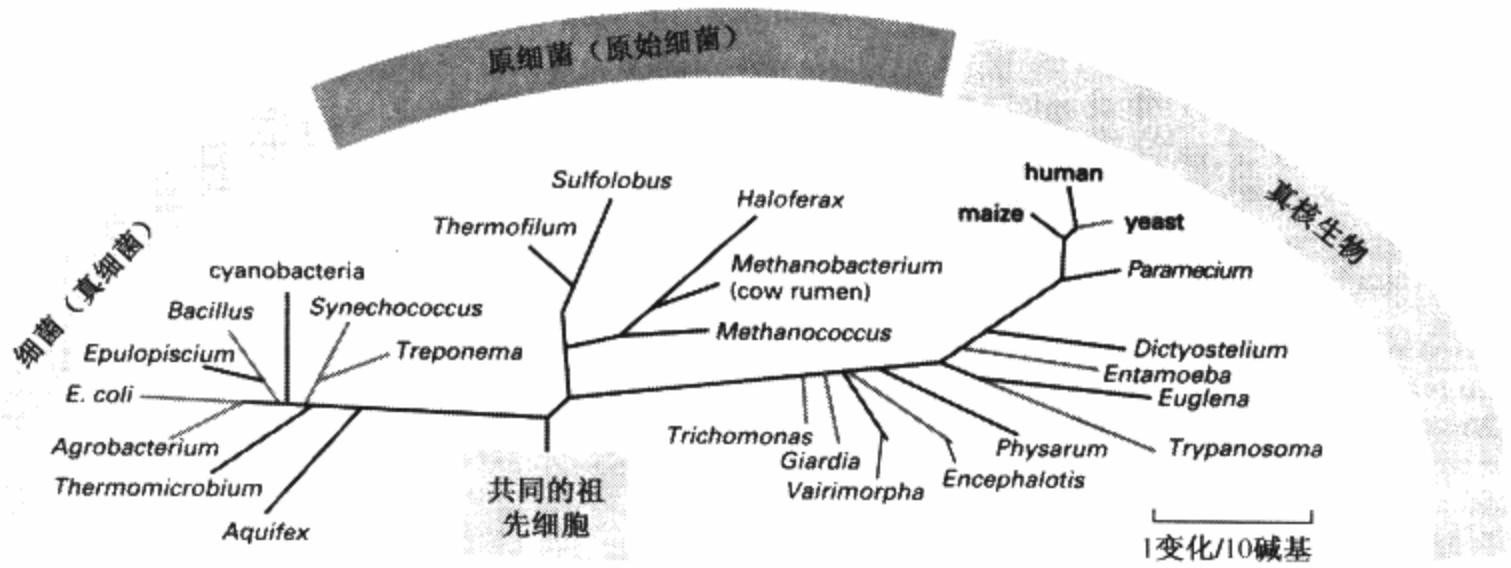


图 1-21　生物界中的三个主要分支

注意，传统上的细菌指的是广义上的原核生物，但最近细菌被重新定义特指真细菌。这也许有点模糊，细菌狭义上特指真细菌。通过比较不同物种的核糖体 RNA 亚基的核酸序列而构建了此树。每个谱系中线的长度代表了此分子进化改变的数目（图 1-22）

古细菌是在人类不愿到达的环境中被首先发现的，如沼泽、污水田、深海、高盐水域以及热酸泉，现在我们知道它们也可存在于在非极端的常见环境中，从土壤、湖泊到家畜的胃中都有它们的存在。在外表上，它们不易和常见的真细菌相区别。在分子水平上，古细菌对遗传信息的处理机制（复制、转录以及翻译）和真核生物相对比较接近，而在代谢和能量转换装置上则和真细菌更接近。我们将在下面的章节里解释这些现象。

一些基因演化迅速，另一些则十分保守

在遗传信息的储存以及复制上，会发生一些随机的事件甚至错误导致核酸序列的改变，即产生**突变**。因此在细胞分裂时，两个子细胞之间以及子细胞和亲代细胞之间都不是完全一样的。这种错误在极少数情况下会引起有利变异，而在大多数情况下是没有任何意义的，在另外一些情况下还会引起严重的损害。例如，可能会打断一个重要蛋白质的编码序列。由于第一种情况改变后的细胞相对而言更可能复制自身，所以这种错误的改变可以永存下来。第二种情况的错误改变（有选择的中性改变）有可能会保存下来，在竞争有限资源的情况下，是突变细胞还是正常细胞胜出只是一个概率事件。而引起严重损伤的突变则会一事无成，突变细胞会在没有子代的情况下死亡。通过这种无休止的突变与自然选择的错误和尝试循环，生物得以进化。遗传特征的改变，赋予它们新的能力，可以更有效地利用环境资源，在与其他个体竞争中获胜，并可以成功地复制自己。

显然，在进化过程中，基因组的一部分会比其他部分更容易发生改变。一部分不编码蛋白质的以及没有特殊调控功能的 DNA 序列的突变是随机的。相反，编码必要蛋白质和 RNA 分子的基因就不是那么容易改变了，当突变发生时，细胞通常会被消灭掉。后一种情况的基因因此而非常保守。经过 35 亿年或者更长时间的进化，基因组中的大部分都已经不是其本来的面目了，但是大多数高度保守的基因在所有的物种中都是相似的。

如果我们要寻找生物树中距离比较远的两个物种间的关系，那么后一种情况的基因就是我们必须要检查的。通过分析一种有 1500 个碱基组成的核糖体 RNA 亚基——16S RNA，我们可以把现有的生物世界分成三大类：真细菌、古细菌以及真核生物。由于翻译的过程对于

生物体来说是最基本的，核糖体的组成自从生命产生以来就是极度保守的（图 1-22）。

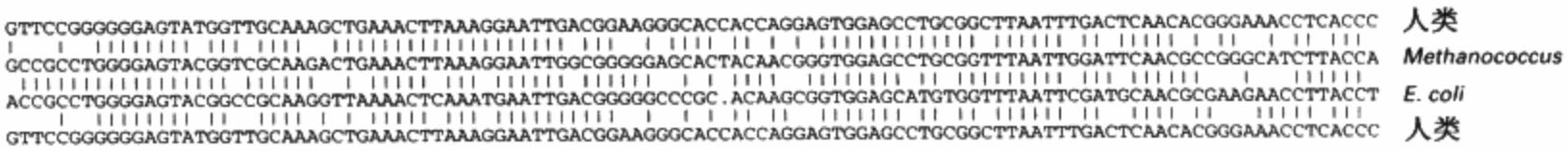


图 1-22　从生命开始遗传信息就是保守的

核糖体两种主要 RNA 成分中较小的一个的部分基因如图所示。平行排列了来自古细菌、真细菌和真核生物中的相应片段；人类的序列下面也排列了一次方便于三种物种两两对比。*E. coli* 序列中间的点表示一个碱基在进化过程中要么在真细菌系中丢失了，要么插入了其他两大分支中。注意来自于 3 个物种的序列代表了生命界中的三大分支，严格来说两两都不相同，但却保留了明显的相似性

大多数细菌和真细菌含有 1000～4000 个基因

自然选择力使原核细胞可以迅速地复制自身，它们从环境中吸取原材料高效地进行复制，对于其复制的最大限制就是食物的供给。小尺寸暗示其面积与体积比巨大，因此可以最大限度地摄入营养物质，提高细胞的复制速率。

人们推测，正是由于上述原因，大多数原核细胞几乎没有携带多余的基因；它们的基因组小而紧凑，基因拥挤在一起并且只有少量的 DNA 调控序列。基因组的小规格使得人们更易于得到它们的 DNA 全序列。人们现在已经掌握了多种真细菌、古细菌以及几种真核生物的全基因组。根据表 1-1 所示，大多数真细菌和古细菌基因组中都有 10^6～10^7 个碱基对，编码 1000～4000 个基因。

表 1-1　一些基因组已全部测序的物种

种类	种类特征	生活环境	基因组大小(每单倍体基因组 1000 核苷酸对)	基因数量(蛋白质)
真细菌				
Mycoplusma genitalium	已知细胞中最小的基因组	人类生殖道	580	468
Synechocystis sp.	光合，造氧(氰基细菌)	湖或溪中	3573	3168
Escherichia coli	实验室的最爱	人类消化道	4639	4289
Helicobacter pylori	导致胃溃疡易致胃癌	人类胃中	1667	1590
Bacillus subtilis	细菌	土壤	4214	4099
Aquifex aeolicus	无机营养的，生活在高温下	火山口	1551	1544
Mycobacterium tuberculosis	导致结核	人体组织	4447	4402
Treponema pallidum	螺旋菌、导致梅毒	人体组织	1138	1041
Rickettsia prowazekii	和线粒体最为相近的细菌，导致斑疹伤寒	虱和人(胞内寄生虫)	1111	834
Thermotoga maritima	有机营养的，生活在高温下	火山口	1860	1877
古细菌				
Methanococcus jannaschii	无机营养、厌氧、产甲烷	火山口	1664	1750
Archaeoglobus fulgidus	无机或有机营养、厌氧、产硫酸盐	火山口	2178	2493
Aeropyrum pernix	好氧、有机营养、热气口	沿海火山	669	2620

续表

种类	种类特征	生活环境	基因组大小(每单倍体基因组 1000 核苷酸对)	基因数量(蛋白质)
真核生物				
Saccharomyces cerevisiae (hudding yeast)	最小的真核模式生物	葡萄皮、啤酒	12,069	~6300
Arabidopsis thaliana (wall cress)	植物模式生物	土壤和空气	~142,000	~26,000
Caenorhabditis elegans (nematode worm)	可预测发育的非常好的简单动物	土壤	~97,000	~19,000
Drosophila melanogaster	对于动物发育遗传学非常重要	腐烂水果	~137,000	~14,000
Homo sapiens (human)	最集中研究的哺乳动物	房子	~3,200,000	~30,000

全 DNA 序列揭示了一个物种中所含有的基因以及它所没有的基因。比较这三大类生物可以发现，哪些基因对它们来说是共同的，由祖先传到当今世界所有的生物体中，而哪些基因是专属于生命树的某一支。要解释这些，我们首先需要更深入地了解基因的起源以及基因组的进化。

新基因产生于先前存在的基因

进化的原材料是已经存在的 DNA 序列，自然界不能产生大段的随机新序列。在这一意义上，没有一个基因是完全意义上的新基因。新的改变可能以下面几种方式产生（图 1-23）：

基因内突变：现有的基因通过内部 DNA 序列的突变而发生改变。

基因重复：现有的基因可以通过复制在同一个细胞内产生一对关系相近的基因。

片段转移：两个或者更多的基因间可以打断并且重连，形成一个含有原本属于不同基因片段的杂合基因。

水平（细胞间）转移：一部分 DNA 可以从一种细胞的基因组转移至另外一种的基因组中，甚至是不同种类的细胞。这种过程和遗传信息由父代转移至子代的垂直转移是相对的。

这些类型的改变都可以在物种 DNA 序列上留下特征性的痕迹，现已有明显的证据表明这 4 种方式都会发生。我们首先集中在其造成的结果上，下面的章节中我们会探讨其发生的机制。

基因复制引起单个细胞内相关基因家族的出现

细胞在分裂为 2 个子细胞时必须复制其全部基因组。但是有时会发生意外使得仅有一部分基因组被复制，这样单个细胞内就会保留有原始的和复制的片段。当这种情况发生时，一个细胞中的 2 个相同基因之一就可以自由突变为执行不同功能的基因。这种复制和分化的过程延续了上百万年，使得基因组中的单个基因产生了一个基因家族。分析原核生物的基因组可以发现许多这种基因家族。例如，在枯草芽孢杆菌中 47%的基因有着一个或多个明显的亲缘基因（图 1-24）。

当基因以上述方式发生复制和分化时，一个物种的个体内就会存在多个来源于同一个原始基因的变异基因。这个进化过程不同于生物物种在生命树分支点分成两个不同后代的遗传差异，如后代的人系与后代的黑猩猩分离。在上述情况下，基因渐渐地在进化过程中产生差异，但它们有可能在两个相近的物种中继续维持相应的功能。基因以这种方式联系，即两个

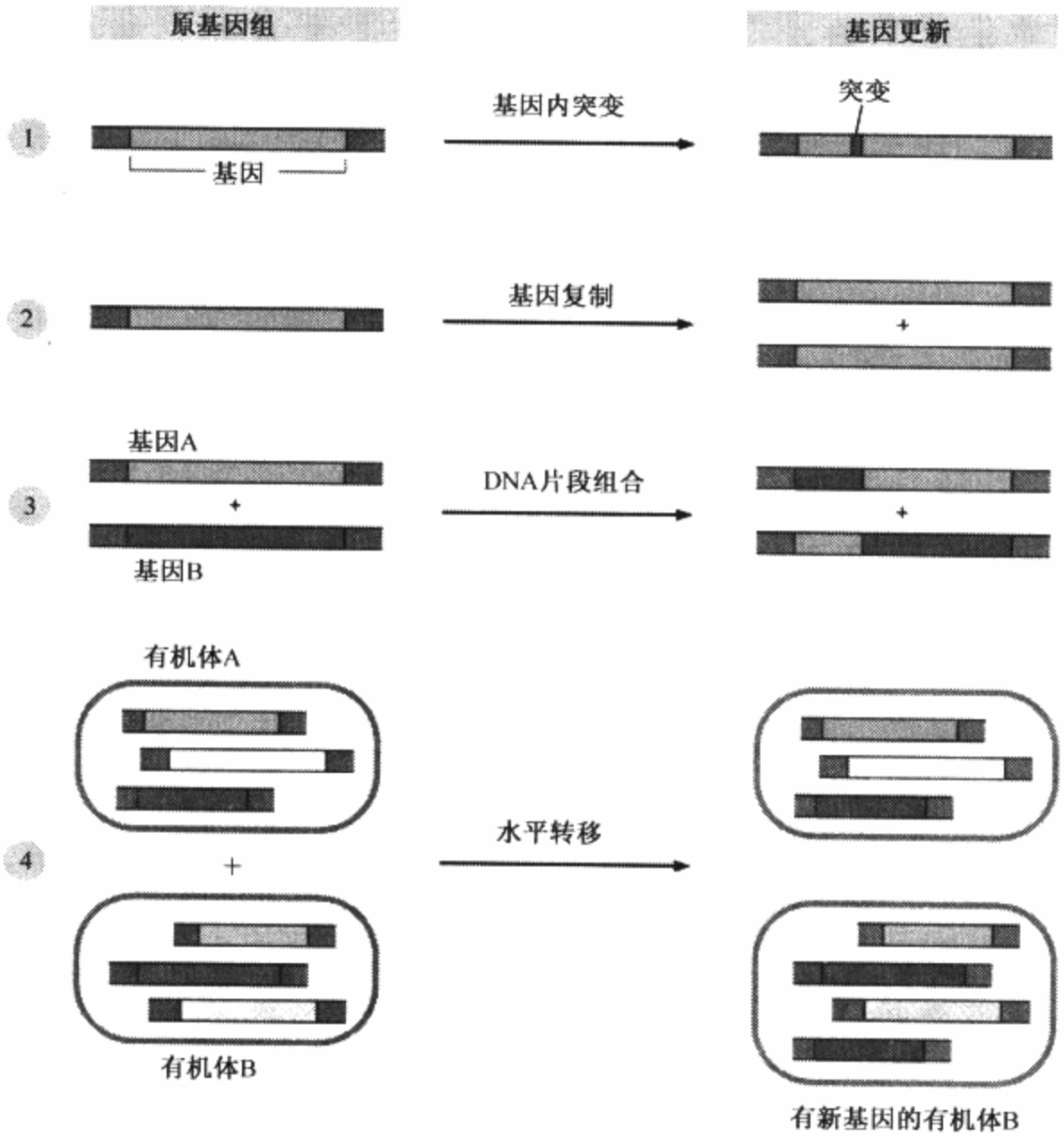


图 1-23　4 种遗传改变的模式和它们对于机体 DNA 序列的影响

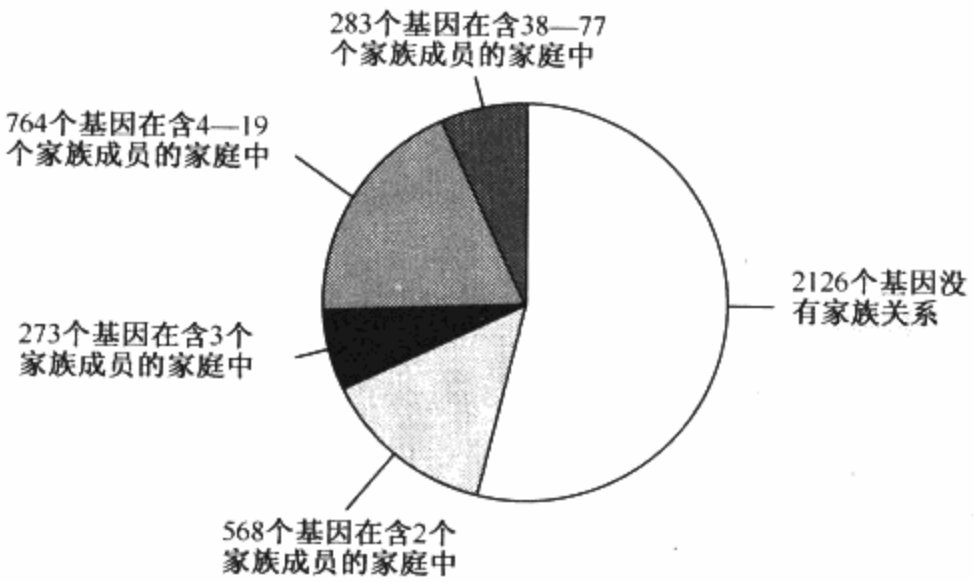


图 1-24　在枯草芽孢杆菌基因组中进化相关的基因家族

其中最大的家族含有 77 个基因，编码 ABC 运载体的多个蛋白质——ABC 运载体是一类存在于三大物种中的膜转运蛋白质（摘自 F. Kunst et al.，*Nature* 390：249～256，1997. © Macmillan Magazines Ltd.）

来源于同一个基因的分化基因最终可以确定两个物种有一个共同祖先，这称为**直系同源**。在单个基因组由于基因复制所导致的相关基因，有可能在功能上产生分化，这称为**旁系同源**。以上述两种方式中的任何一种相互关联的基因都称之为**同源基因**，它是一个包含两种关系的

通用术语（图 1-25）。

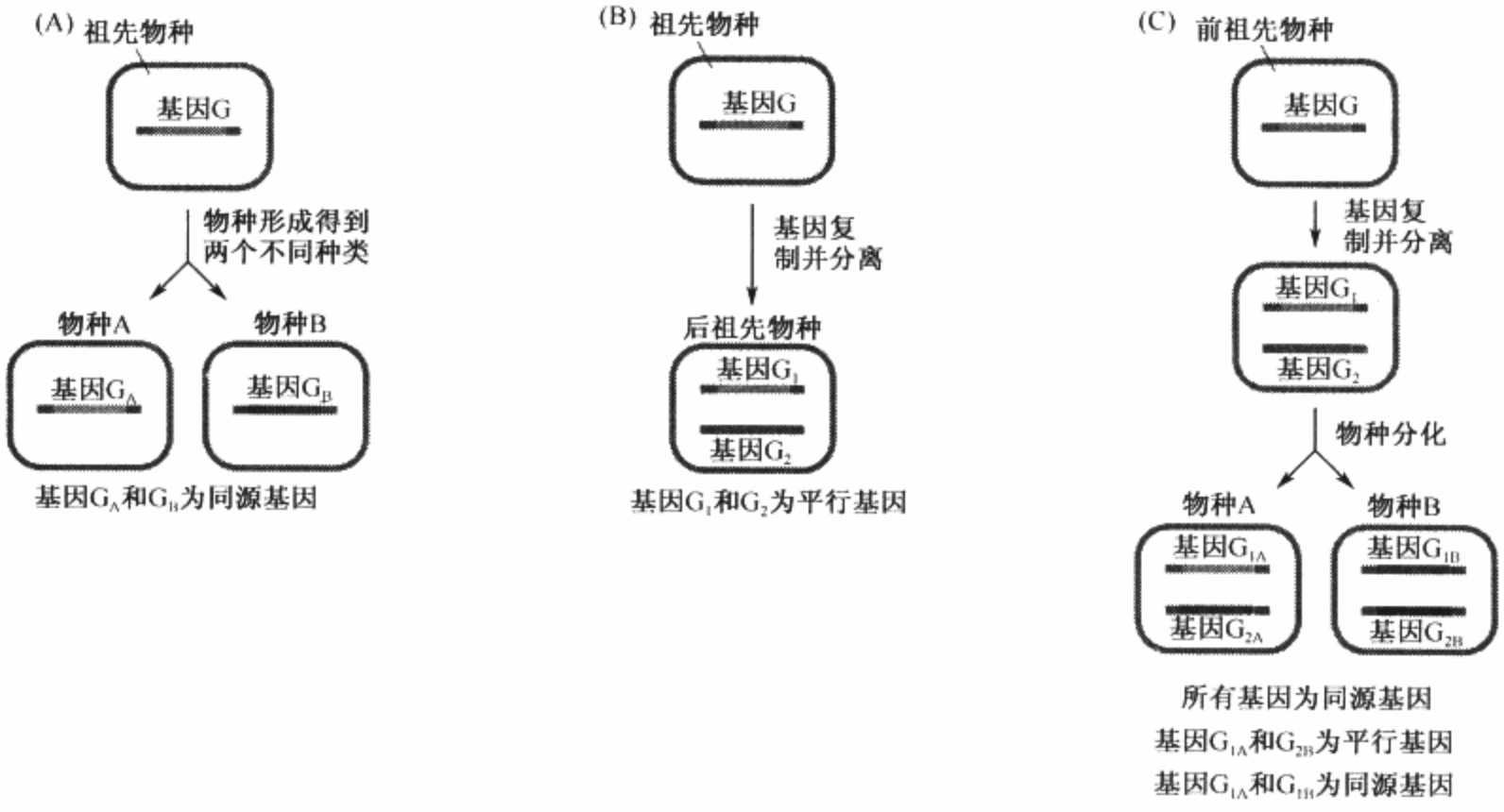


图 1-25　平行基因和同源基因：有着不同进化途径的两种基因同源性

（A）和（B）最基本的可能性。（C）可能发生的复杂机制

基因之间的家族谱系关系可变得非常复杂（图 1-26）。例如，一个拥有旁系同源家族的生物体（如血红蛋白基因 α、β、γ、δ、ε、ζ 和 θ）进化成两个独立的物种（如人和黑猩猩），每个物种都拥有一套完整的旁系同源组。所有的 14 个基因是同源物，人的血红蛋白 α 和猩猩的血红蛋白 α 是直系同源，而人或猩猩的血红蛋白 β 是旁系同源等。并且，脊椎动物的血红蛋白（血液中的氧结合蛋白）与脊椎动物的肌红蛋白（肌肉中的氧结合蛋白）同源，以及与脊椎动物和植物中编码氧结合蛋白的更远的基因同源。一般很容易从根据 DNA 序列识别两个不同物种基因同源；在没有其他信息的情况下，更难决定它们是否是直系同源。

基因可以在两个物种之间相互转移，这种现象在实验室和自然界都可以发生

有证据表明原核生物中的基因可以在两种物种中转移。最明显的是细菌**病毒**，又叫做**噬菌体**的 DNA 序列（图 1-27）。这种遗传材料的小包裹体进化成可以利用宿主细胞内的复制和生化合成途径的寄生虫。它们在一种细胞内复制，又以一种防护的包装物形式从细胞中释放，然后去浸染另一个同类或非同类细胞。它们在细胞内能以一种独立的 DNA 片段存在，叫做质粒，也可以插入到宿主基因组中。在此过程中，病毒有时可以携带一部分宿主 DNA 并将其运送至另外一种细胞的基因组中。这种遗传物质的传递在原核生物中经常发生，在同种真核细胞中也时常存在。

水平转移在不同物种的真核细胞中并不常见，并且这种转移好像在真核生物的进化中没有起到明显作用。相反，不同种类的真核生物中垂直基因传递却经常发生。许多真核生物拥有不平常的能力，可以从周身环境中吸取并非病毒 DNA 的分子，因而捕获了许多外源分子所携带的遗传信息。这使得野生细菌易于从相关的临近细胞中获取基因。例如，抵御抗生素

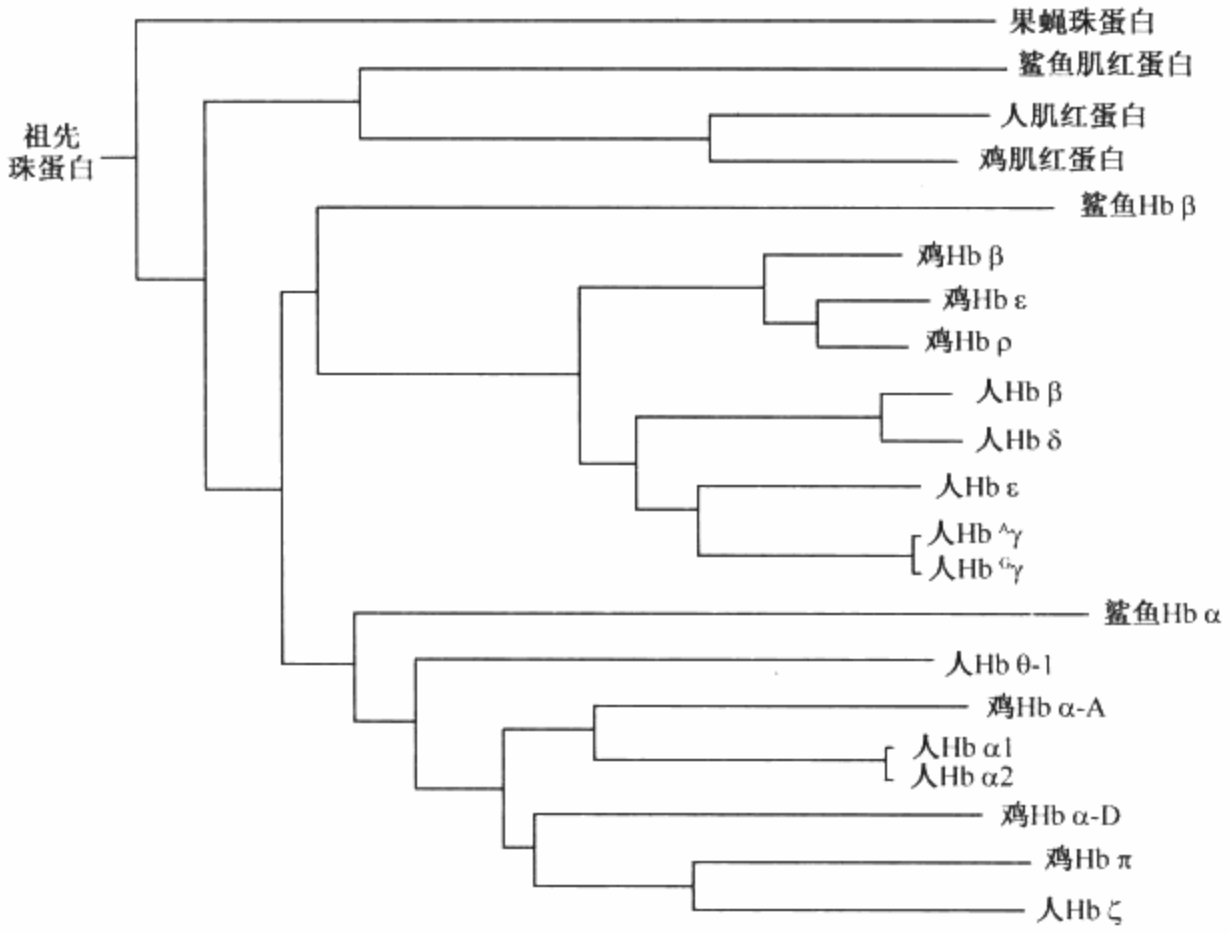


图 1-26　同源基因的复杂家族

此图显示了人、鸡、鲨鱼和果蝇中血红素（Hb），肌红蛋白质以及珠蛋白的谱系关系。谱线的长度代表了氨基酸序列分歧的数量

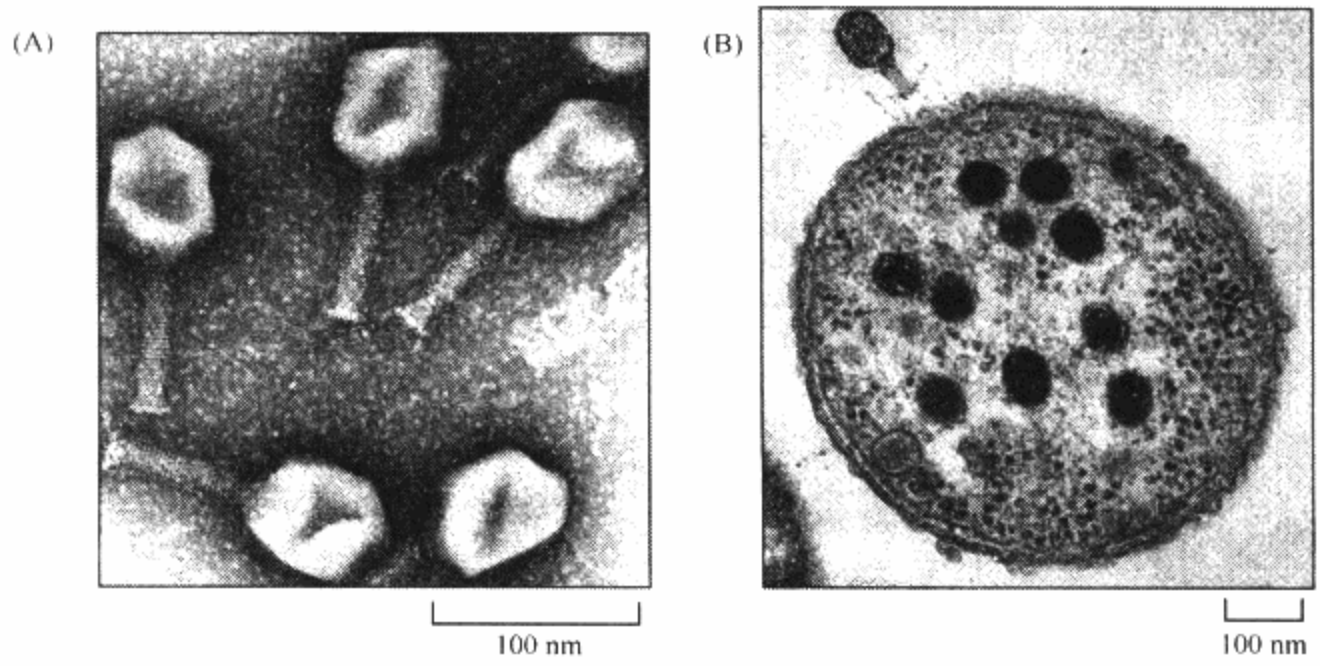


图 1-27　DNA 通过病毒从一个细胞转移到另一个细胞

（A）一种细菌病毒颗粒，T4 噬菌体的电镜照片。病毒的头部含有病毒 DNA，尾部带有将 DNA 注射到宿主细胞的装置。（B）T4 噬菌体附着细菌表面的横切面。细菌体内巨大的黑色物是正在装配的新 T4 病毒颗粒的头部。当它们成熟后，细菌就会破裂从而释放出病毒［（A）摘自 James Paulson；（B）摘自 Jonathan King and Erika Hartwig from G. Karp, Cell and Molecular Biology, 2nd edn. New York: John Wiley & Sons, 1999. © John Wiley.］

或者防止毒性物质产生的基因可以由一个物种转移至另一物种，从而赋予受体菌一种选择优势。人们发现，细菌可以在其生态系统中以这种方式产生新的甚至是危险的菌株，它们存在于医院中还有人体中适当的部位。例如，这种基因水平转移的方式产生过一种青霉素抗性的奈瑟淋病菌，这种病菌在过去的 40 年中造成了淋病的广泛蔓延。如果扩大时间范围来看，这种基因水平转移的方式更是影响重大，据估计，在过去的 1 亿年中，*E. coli* 基因组中 18%的

基因是从其他物种中获得的。

物种中基因信息的水平交换是由性引起的

目前，遗传信息的水平交换在细菌进化中起重要作用，这种交换有可能在生命出现初期发生得更为频繁和混乱。实际上，人们推测，现今真细菌、原细菌以及真核生物的基因组并不是由单个类型的祖先基因组中分化而来的，而是由不同细胞中的三种相互独立的基因原始集合频繁交流所产生的（图1-28）。这就可以解释那个令人难解的现象，即真核生物在遗传信息处理的基本方式上（如 DNA 复制、转录和翻译中）和古细菌更相似，而在新陈代谢的步骤上和真细菌更相似。

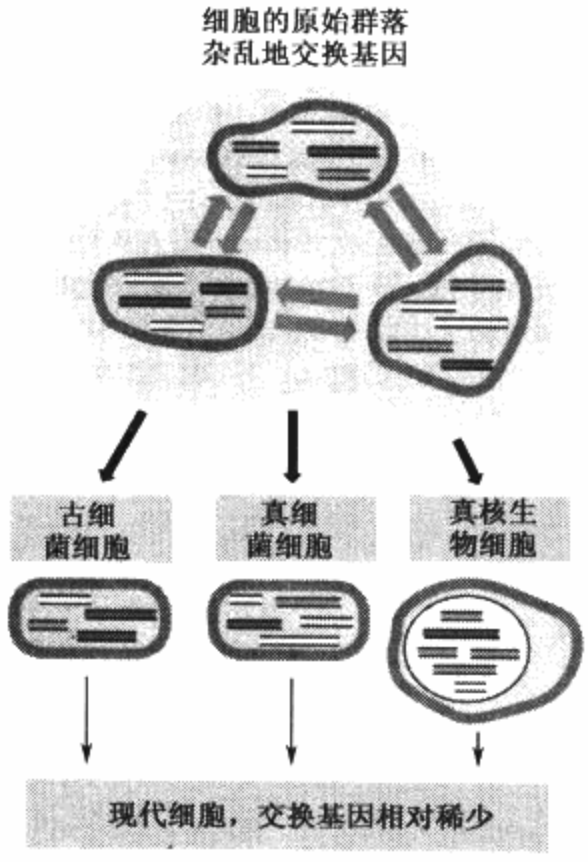


图 1-28　早期进化中的水平基因转移

在地球生命的早期，细胞较难保持它们独立的特征，基因交换较现在来说也比较频繁。这样来说，古细菌、真细菌以及真核生物三大谱系可能由原始杂乱交换基因的细胞群体遗传了不同的但是有重叠亚群的基因

基因在细菌中可以水平转移好像是一个令人惊奇的步骤，但它们有一个跟我们人类类似的现象：性。有性生殖会引起遗传信息在两个初始的细胞系间进行大规模的平移——这两个初始的细胞系即父细胞系和母细胞系。当然，性的一个主要特征是遗传交换通常发生在一类物种的两个个体间。但不论是同物种或是不同物种，水平基因转移会留下一个典型的特征：产生的个体在其一部分基因上和双亲间的一个接近，而另一部分和双亲间的另一个接近。如果外星智慧生物能够比较人类个体的基因组，即使他们不知道人类的行为，也会推论出人类是进行有性生殖的。

有性生殖是一个广泛的（虽然不是普遍的）现象，尤其是在真核生物中。细菌甚至时时放纵于和自身物种的其他个体进行有性交换 DNA 中。自然选择显然赋予了物种这种行为的能力，尽管进化理论学家仍在争论性的选择优势是什么。

基因的功能常能根据其序列而推测

基因的家族关系不仅对于其历史研究有着重要意义，而且大大简化了解密基因功能的工作。一种新的基因序列一旦被确定，就有可能只通过在计算机上敲打几个简单的字母，在含有已知基因序列的数据库中检索和其相关的基因。许多情况下，一种或者多种这种同源基因的功能已经被试验证明了，接下来由于基因序列决定其功能，人们可以预测新基因的功能，有可能是和已知的同源基因功能相同。

按照这种方式，我们可以通过分析某种生物基因组的 DNA 序列，并且利用已经深入了解的其他物种的基因功能信息，来预测其大量的生物学特征。结核分枝杆菌——一种引起结核病的细菌，在实验室中极难通过实验体系进行研究，但为强有力的比较基因组学提供了一个例子。DNA 测序显示其基因组大小为 4 411 529 个碱基对，包含大约 4000 个基因。在这些基因中，人们发现（在 1998 年全基因序列测出时）有 40%的基因和其他物种的已知基因同源，因而可以在此基础上推测其功能。另外有 44%的基因在信息学上和其他基因相似。例如，和某个较长的蛋白质序列有一个保守的蛋白质结构域。4000 个基因中只有 16%是人们所

不熟悉的。另外枯草芽孢杆菌（图 1-24）序列显示有将近一半的基因和结核分枝杆菌非常相似，表示它们一定都起源于相关基因的复制。和其他细菌相比，结核分枝杆菌中存在大量编码脂肪分子合成和降解过程中的酶类的基因。这些结果表明，此种菌的不平常外壳有可能富含这些底物，外壳以及产生它的酶类可以用来解释结核分枝杆菌是怎样逃避结核病患者免疫系统的。

生命树上三个基本分支间有 200 多个基因家族相同

给定三种类型生物（原细菌、古细菌以及真核生物）各自的代表生物的全基因组序列，人们可以系统地寻找可以跨越这种巨大进化分支的同源基因。按照这种方式我们可以观察所有物种一般的遗传方式。但这种方式中也存在巨大的困难。例如，某个物种可能丢失了祖先的一部分基因，而它所含有的另外一些基因可能是由于水平转移而获得的，并不是祖先遗传下来的。最近的基因组比对显示，至少在原核世界中，种系特异性基因丢失和水平基因转移都是进化的两个重要因子，甚至存在于两个进化上相距较远的物种中。最后，经过 20 亿～30 亿年的进化过程，最初的基因构成发生了变化，以至用我们目前的方法无法进行识别。

由于进化过程中的这些奇特行为，目前好像只能辨认一小部分原始基因。因而在比较 18 种真细菌、6 种古细菌和 1 种真核生物（酵母）的 2264 个编码基因后，仅有 76 种是真正相同的（即在所有分析物种中都存在）。这些相同家族中的大部分是转录、翻译系统的组分，这些并不像是祖先基因的现实世界替代物。一种稍好的方法（虽然依旧粗略）是从三大类物种中的多种生物中计算其共有的基因家族。大部分的基因家族都有了相应的功能（至少是大致的生化功能，但大多都是预测的），其中大部分的用于转录、核糖体产生、氨基酸新陈代谢以及运输等步骤（表 1-2）。这一部分高度保守的基因家族仅仅反映了所有模式生物共同遗传性的大体框架，而构建出共同祖先的原始基因图需要进一步的基因组测序以及更为细致的分析手段。

表 1-2　目前生物界三大物种相同基因家族的数量及功能分类

基因家族功能	“普通”家族的数量
翻译，核糖体结构和生物发生	61
转录	5
复制，修复，重组	13
细胞分裂及染色体分区	1
分子伴侣	9
外膜，细胞壁生物发生	3
分泌	4
无机铁转运	9
信号转导	1
能量产生和转换	18
碳水化合物代谢及转运	14
氨基酸代谢及转运	40
核酸代谢及转运	15
辅酶代谢	23
脂类代谢	8
预计的广义生物化学功能；未知的特殊生物学功能	33
未知功能	1

注：为完全分析，把在至少两种不同古细菌（*Archaeoglobus fulgidus* 和 *Aeropyrum pernix*）两种进化上远离的细菌（*Escherichia coli* 和 *Bacillus*）以及一种真核生物（酵母，*Saccharomyces cerevisiae*）都具代表性的基因家族定义为“普遍”（数据来自 R. L. Tatusov, E. V. Koonin, and D. J. Lipman, *Science* 278: 631-637, 1997; R. L. Tatusov, M. Y. Galperin, D. A. Natale, and E. V. Koonin, *Nucleic Acids Res.* 28: 33-36, 2000; and E. V. Koonin, personal communication.）

突变揭示基因的功能

仅仅专注于基因组序列而没有更多信息的情况下，我们并不能揭示基因的功能。例如，我们可以看出基因 B 和基因 A 相似，然而我们首先要如何才能知道基因 A 的功能呢？即使我们了解了基因 A 的功能，我们又如何确定 B 的功能是否跟其序列显示的那样和 A 相似呢？我们又怎样才能把遗传世界和现实生物世界联系起来呢？

基因功能的分析极大地依赖于两种互补的方法：遗传学和生物化学。遗传学主要是以研究突变体为主，我们可以找出或者制造基因改变的生物体，并且研究基因的改变对生物体结构以及功能的影响（图 1-29）。生物化学方法主要研究分子的功能，我们从机体中提取出感兴趣分子，来研究其化学功能。通过生物化学和遗传学方法相联系，我们可以观察一种突变种体内的化学变化，并可能找出依赖于特定基因的分子。同时，研究突变体的功能可以让我们从总体上了解这些分子在机体中的作用。因此，遗传学和生物化学方法的相联合为人们研究生物体基因、分子、结构以及功能间的联系提供了重要途径。

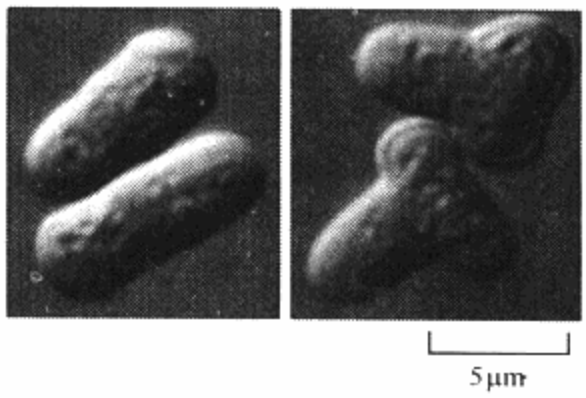


图 1-29　突变的表型反应了基因的功能

正常酵母细胞和突变了一个基因的突变细胞的比较，突变细胞由雪茄形（左）变成了 T 形（右）。所以突变的基因具有控制细胞形态的功能。但目前从分子水平来说，是否是基因的产物行使了此项功能呢？这是一个比较难的问题，需要生化学家来解答（经 Kenneth Sawin 和 Paul Nurse 许可）

近年来，DNA 序列信息和分子生物学的有力工具发展迅速，通过序列比对，人们常常在一个基因内发现进化过程中保守的结构域。这些保守结构域可能是决定其功能的重要部分，我们可以在实验室内对某个基因的特殊位点进行突变，或人为融合两个不同基因的部分序列，从而验证它对于基因产物的影响。

生物体受基因的驱动产生大量的特定 RNA 以及蛋白质，以利于后续的生化合成。分子结构学者们可以确定基因产物的三维构象，从而揭示每个分子在其中的位置。生化学家们可以确定遗传特定分子中的每一部分对于其功能起到什么样的作用。细胞生物学家们则可分析细胞受到驱动表达突变基因的方式。

但是，基因的功能不能靠一种方法来确定，也没有一种简单的标准模式可以描述其功能。例如，我们可以发现某种基因产物可以催化某种化学反应，但我们对于这种反应对于机体的重要性却没有概念。每当一种新基因家族成员的功能并不像它们的序列所显示的那样时，我们的生物学家就会面临一个新的挑战。而且一个基因的功能并不能全面的解密，除非我们从整体上了解了其在生物或器官中的功能。因此如果我们最终要确定基因的功能。我们必须研究整个机体而并不是仅限于分子或者细胞水平。

分子生物学家将焦点对准了大肠杆菌

由于生物体非常复杂，所以我们对于一种生物体了解越多，这种生物体就越会成为我们下一步研究的目标。每个发现都会提出新的问题，但也会为在这种生物体内解决这些问题提出新的工具。因此，同一种模式生物不同特征的揭示要归功于研究它的庞大生物学家群体。

在形形色色的细菌世界中，分子生物学的焦点很长时间内都集中在一种生物——大肠杆菌上（图 1-17 和图 1-18）。这种杆状的小型真细菌细胞生活在人类或其他脊椎动物肠道中，

但它也可以很容易地在培养瓶中的营养肉汤中生长。进化赋予了它可以适应多种化学环境以及繁殖迅速的特点。大肠杆菌的遗传指令存在于一个单链环状的 DNA 分子中，有 4 639 221 个碱基对，可以编码大约 4300 个不同的蛋白质（图 1-30）。

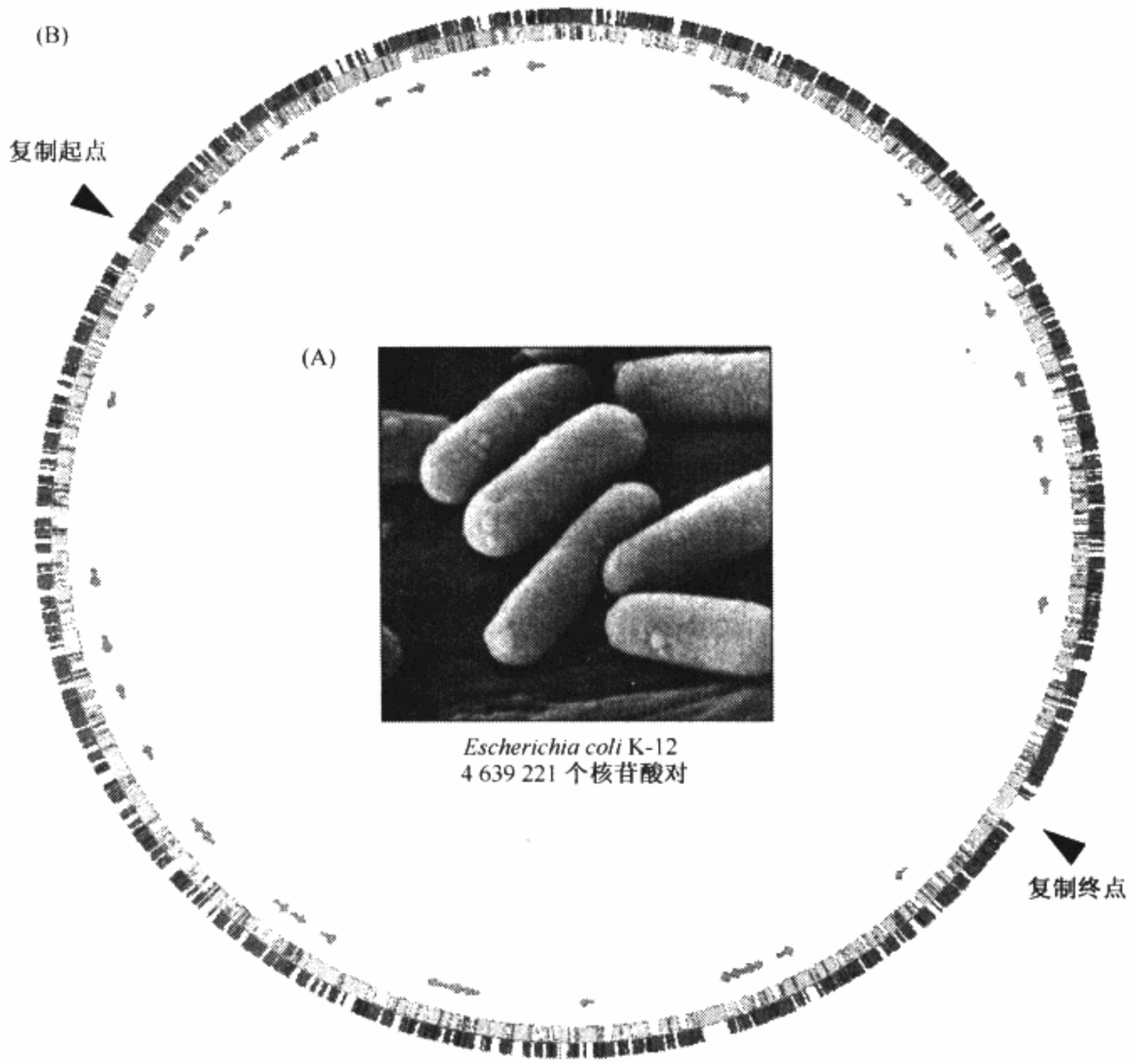


图 1-30　*E. coli* 基因组

（A）*E. coli* 细胞的集群。（B）*E. coli* 基因组中 4 639 221 个核苷酸对的示意图（*E. coli* K-12）。示意图是环形的，由于 *E. coli* 的 DNA 和其他原核生物一样形成一个封闭的单链环。编码蛋白质的基因以黄色或橘色条表示，仅编码 RNA 的基因由绿色箭头表示。一些基因由 DNA 双螺旋的一条链转录（以顺时针方向表示），其他的是另一条链所转录（逆时针方向）［（A）摘自 Tony Brain and the Science Photo Library；（B）F. R. Blattner et al., *Science* 277：1453～1462，1997. ©AAAS.］

从分子角度讲，大肠杆菌是我们了解最清楚的物种。我们所了解的生命的基本机制，例如细胞怎样把自身的 DNA 传递给子代，怎样解码 DNA 中含有的指令以知道特定蛋白质的合成，都是来源于 *E. coli* 的研究。在进化过程中，基本的遗传机制都是高度保守的，因此 *E. coli* 体内的这种机制本质上和我们体内的是相同的。

小结

原核生物（细胞没有明显细胞核）在生化上是一类种类最为广泛的生物，包括可以从无机化学环境中，如由大洋底热岩浆中含有的矿石反应物中，吸取能量和营养的物种——35 亿年前的第一个细胞有可能就是以此为营养。DNA 序列比对揭示了生物之间的联系，表明原核

细胞在进化过程中分成了两类：细菌（或真细菌）和古细菌。它们和真核生物（细胞有一个膜包裹的细胞核）一起组成了生物树的3个基本分支。大多数细菌和古细菌都是小的单细胞生物，有着紧凑的基因组，1000～4000个基因。一个物种的许多基因的DNA序列有着很强的家族相似性，显示它们可能由于基因复制分离而起源于同一个祖先基因。家族相似性（同源性）在不同物种的基因序列比对中都非常明显，有200个基因家族高度保守，被公认是所有三大类生物共有的。因此，如果有一个新发现的基因序列，人们可以根据它在深入研究的模式生物，如大肠杆菌中的同源基因来推测出其功能。

真核生物的遗传信息

总体来说，真核细胞相对于原核细胞来说，体积更大，内部更精细，基因组也更大更精细。更大的规格伴随着的是细胞结构和功能上的根本不同。而且大多数真核细胞形成多细胞组织，达到一个原核细胞无法达到的复杂性。

由于真核细胞非常复杂，使得分子生物学家遇到一系列的挑战，我们在此书的其他部分会讨论到这一点。生物学家们越来越多的通过分析及操作细胞和机体内的遗传信息来应付这种挑战。因此首先了解一些真核细胞基因组的特殊特征是非常重要的。我们首先简要的叙述真核细胞是如何构造的，怎样反映出它们的生活方式以及它们的基因组和原核生物有什么不同。接着是分子生物学家利用遗传信息，尝试发现真核生物运行机制的策略方法。

真核生物起源于捕食生物

真核生物的定义就是可以把其DNA放于一个内部独立空间——细胞核的细胞。核膜是

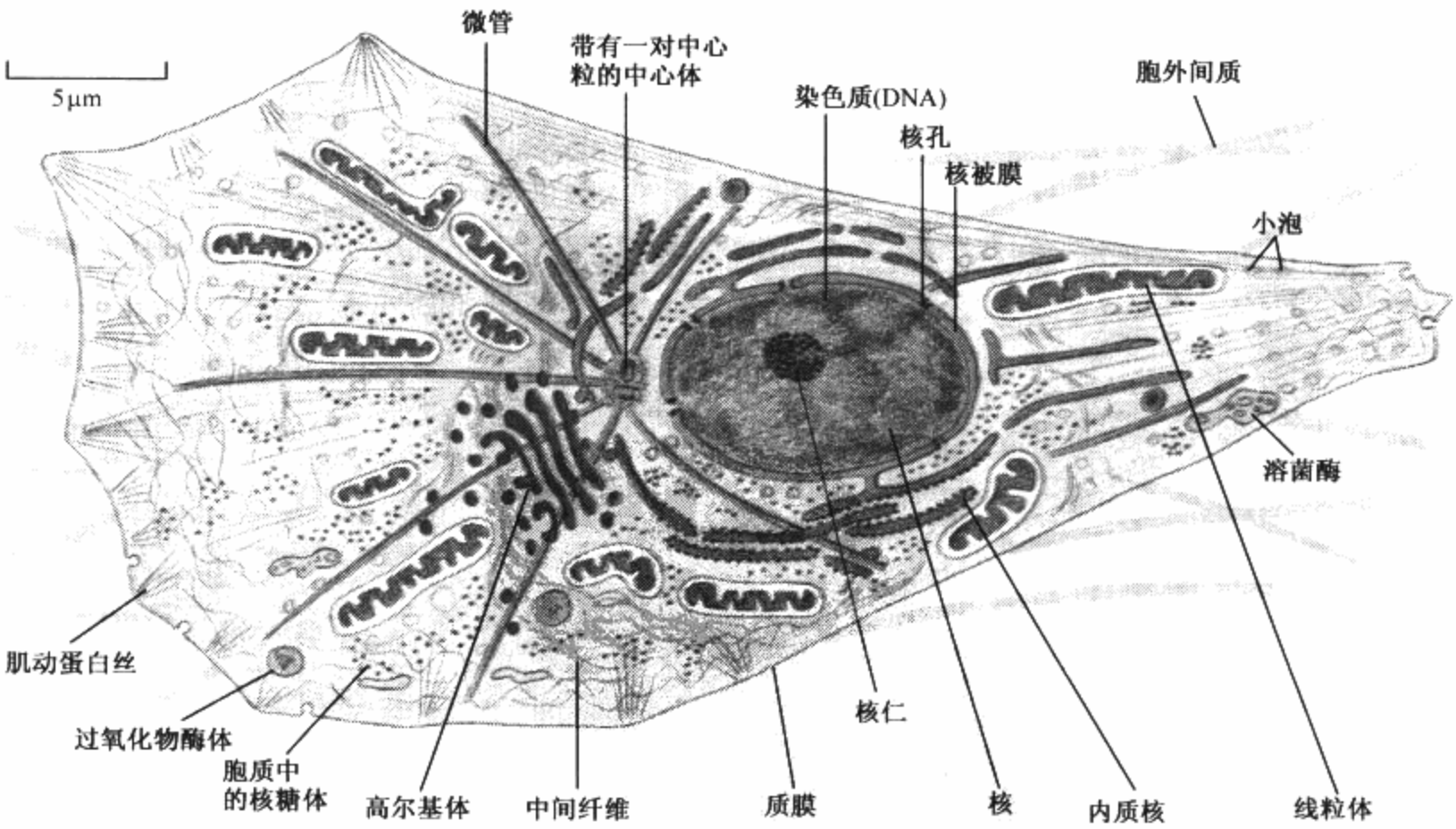


图1-31　真核细胞的主要特征

此图描绘了一个典型的动物细胞，但几乎所有的组分都存在于植物、真菌以及单细胞真核细胞如酵母和原生动物中。除了图上显示的组分，植物细胞中还含有叶绿体，并且细胞膜外包被有一层纤维素组成的坚硬胞壁

一个双层膜结构，把 DNA 和细胞质分开。真核生物还有和原核生物相区分的其他特点（图 1-31）。一般而言，它们的细胞长度是原核生物的 10 倍，体积则有 1000 倍。它们有细胞骨架，一个由蛋白质微管交叉于细胞质中形成的系统，上面附着有多种蛋白质，在细胞中可作为桥梁、绳索、细胞机械强度的动力，可以控制细胞外形，驱动并且导向其运动。核膜只是一系列复杂内膜的一部分，每种内膜结构上都跟细胞膜相似，只是包裹细胞内的不同空间，其中许多在消化和分泌有关的步骤中发挥作用。缺少了原核生物的粗糙细胞壁，动物细胞和自由生活真核细胞原虫可以迅速改变自身外形，并且可以通过吞噬作用（图 1-32）吸入其他细胞或小物体。

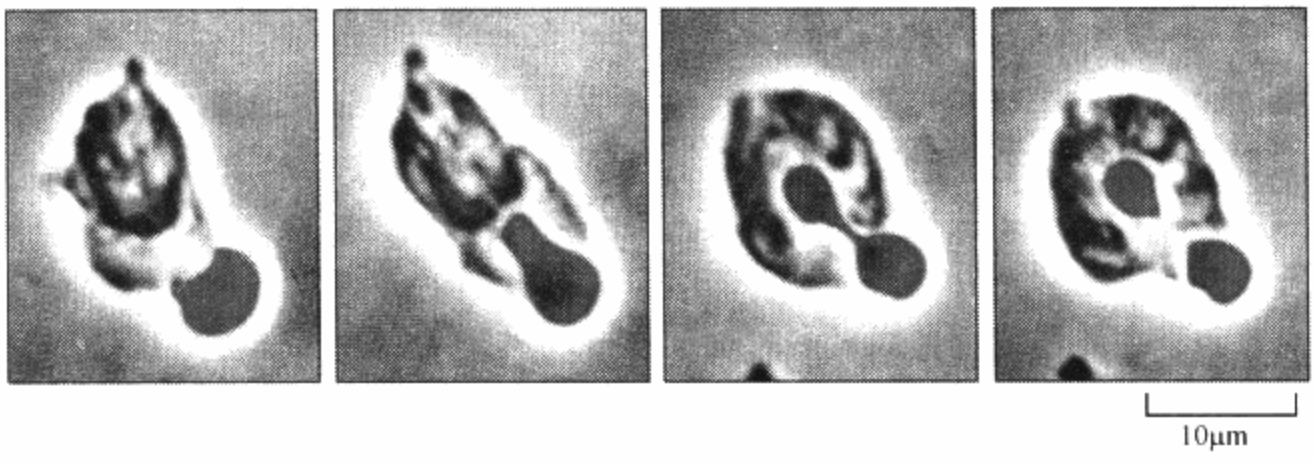


图 1-32　吞噬作用

这一幅幅静止的画面可以组成一部演示人类白细胞（中性白细胞）吞食抗体处理后的红细胞（红色标记）的电影（经 Stephen E. Malawista 和 Anne de Boisfleury Chevance 许可）

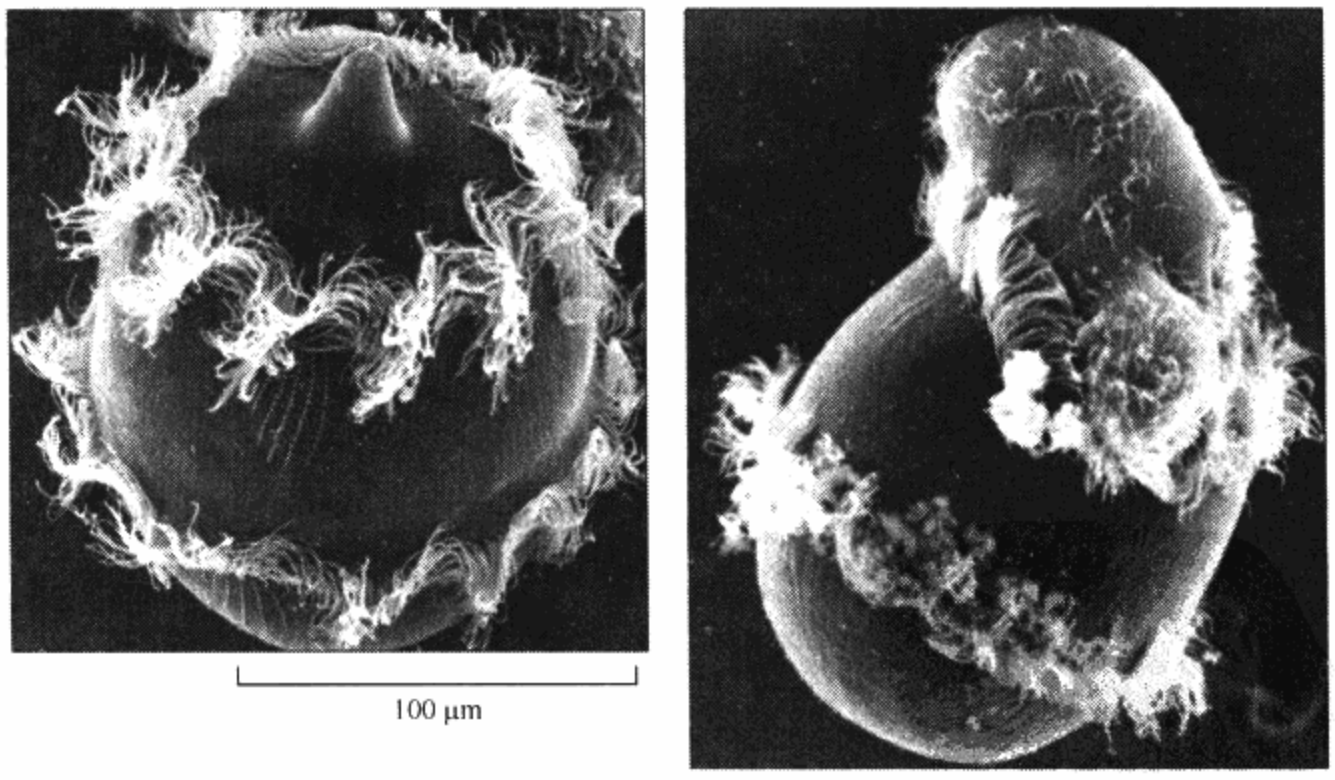


图 1-33　吞食其他细胞的单细胞真核生物

（A）*Didinium* 是一种肉食性原生动物，属于纤毛虫的一种。其细胞呈球形，直径大约 150μm，边缘环绕两层纤毛——弯曲的鞭状纤毛不停地摆动；除了一个突出的像鼻子一样的器官外前端平滑。（B）*Didinium* 通常藉于纤毛的同步摆动而在水中高速运动。当遇到合适的食物时（通常是其他种类的原生动物）它就会由鼻状物中释放出大量的有麻痹作用的小分子。然后 *Didinium* 附着其上并通过吞噬作用吞噬其他细胞，就像一个空心球一样翻转过来包住跟自己体积差不多大的食物（经 D. Barlow 许可）

这些特征究竟是怎么演变的，在哪些序列上有变化，至今仍是个谜。一种合理的解释是两种细胞都反应了生命生存的方式，真核细胞起源于捕食者，通过捕食其他细胞得以生存（图 1-33）。这样一种生存方式需要一个有着柔韧细胞壁的庞大细胞和一个可以支持、指导细胞膜移动的精细骨架。同时需要细胞内长而易断的 DNA 分子隐藏于独立的细胞核中，以避免细胞骨架移动过程中对基因组造成的伤害。

真核细胞起源于共生体

捕食的生存方式也可以解释一些真核细胞的其他特征。几乎所有的真核细胞都含有线粒体（图 1-34）。线粒体是细胞质中的一种小体，外有双层膜包裹，可以由食物分子，如糖类的氧化中吸取氧气和能量，以产生驱动细胞活动的 ATP。线粒体大小和小细菌差不多，并且像细菌一样拥有自己环状 DNA 的基因组，有着和真核细胞不同的核糖体，以及自身的转运 RNA。目前可以确定线粒体起源于自由生长的有氧代谢（好氧性的）细菌，后来被远古的不能利用氧气的（厌氧性的）真核细胞所吞噬。被吞食的细菌没有被消化，反而进化出和真核细胞及其后代共生的特征，从宿主这里寻求保护及营养，并且供给宿主能量（图 1-35）。这种原始厌氧捕食性细胞和好氧细菌细胞的共生关系预计在 15 亿年以前就建立起来了，当时的地球大气刚刚富集了氧气。

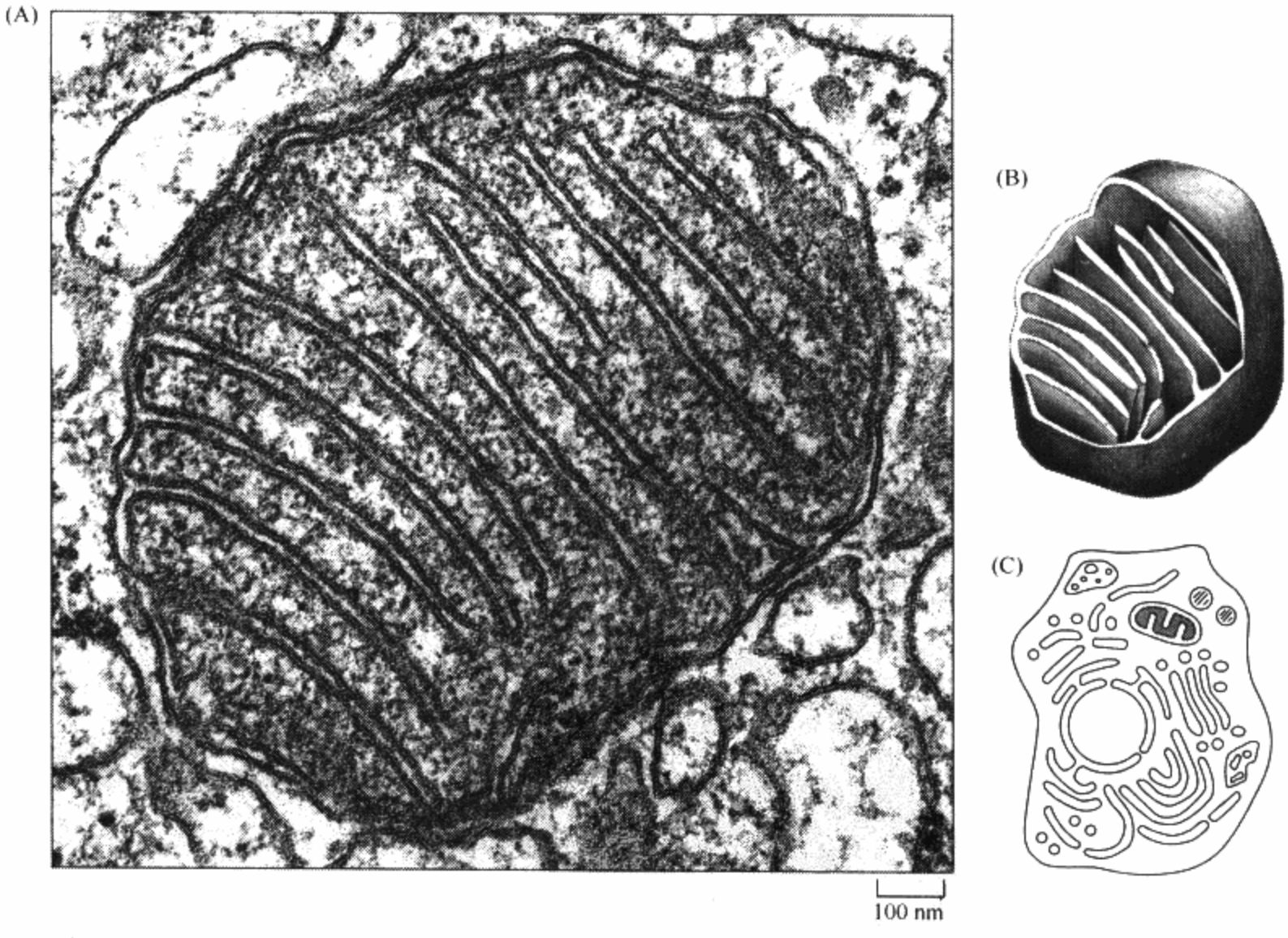


图 1-34　线粒体

（A）电镜下的横切面。（B）线粒体的简图，显示了其三维结构。（C）真核细胞简图，内部的线粒体包含其自己的 DNA 和核糖体，已用颜色标出。注意其光滑的外表膜和卷曲的内表膜，内部包裹了可以由食物分子的氧化中产生 ATP 的蛋白质［（A）经 Daniel S. Friend 许可］

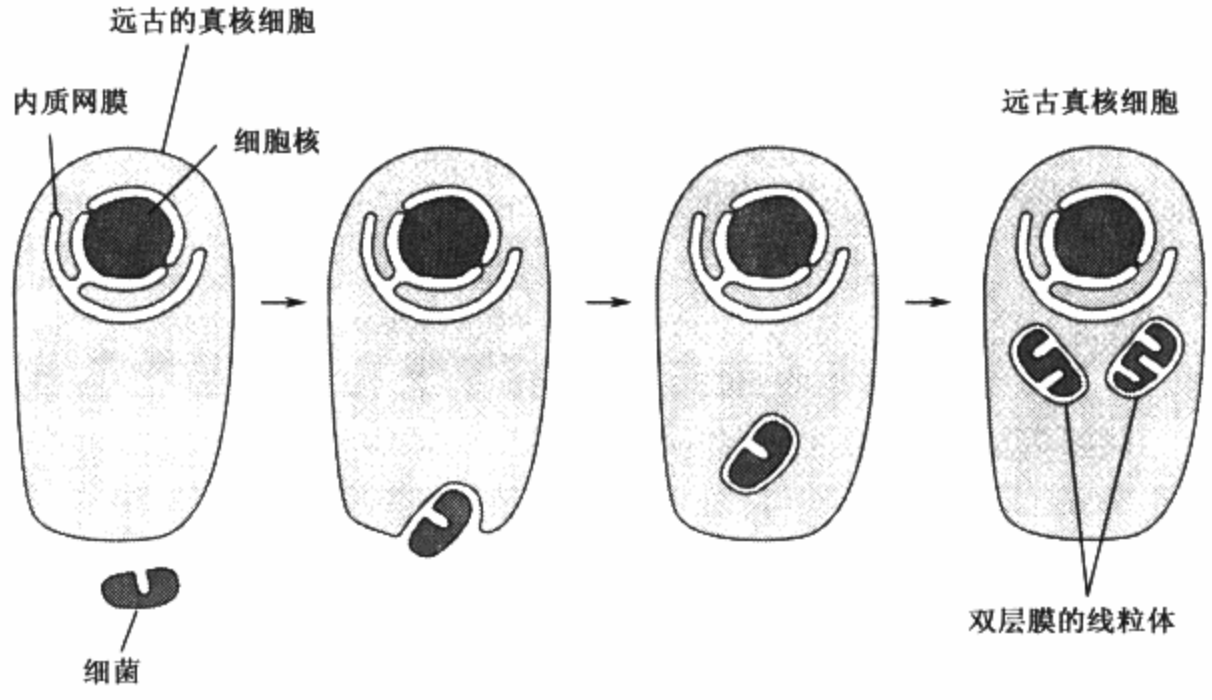


图 1-35　线粒体的起源

预测远古的真核细胞吞噬了一种远古的细菌，即线粒体的祖先，形成了一种共生的关系

许多真核细胞（如植物和藻类）还含有其他一种和线粒体有些类似的有膜包围的小细胞器——叶绿体（图 1-36）。叶绿体执行光合作用，可以利用太阳能由大气二氧化碳和水中合成碳水化合物，然后把产物作为食物传递给宿主细胞。和线粒体一样，叶绿体也含有自身的基因组，可以确定是含有线粒体的细胞捕获了光合细菌并形成了共生体系（图 1-37）。

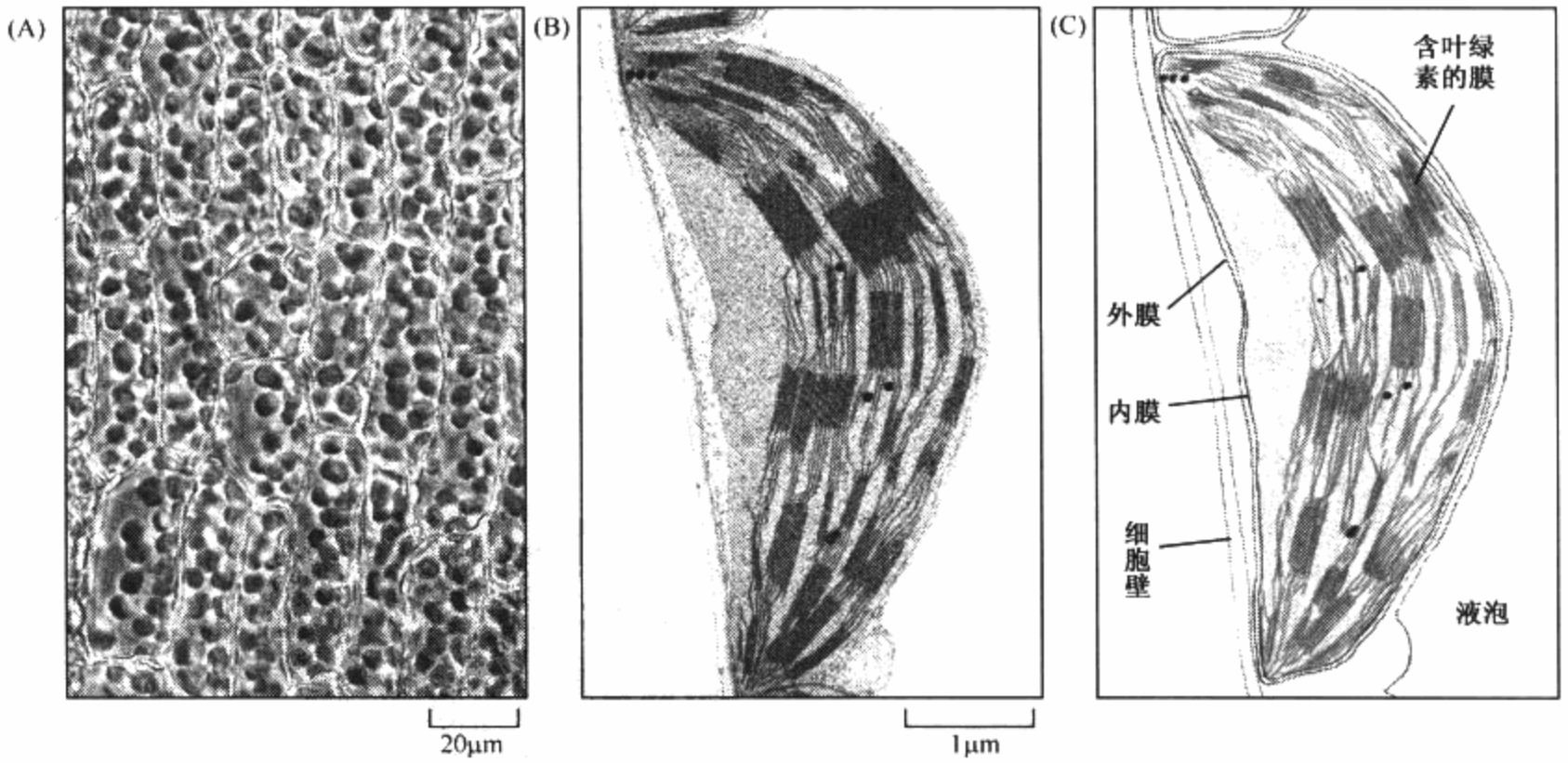


图 1-36　叶绿体

在植物细胞和一些单细胞真核生物中存在这种可以利用太阳能的器官。（A）光学显微镜下的苔藓叶细胞，绿色的叶绿体可见。（B）草叶中叶绿体的电镜照片，其内膜是高度折叠的，内部含有可以吸收太阳光的叶绿素分子。（C）为（B）的草图［（B）经 Eldon Newcomb 许可］

有了叶绿体的真核细胞不需要以其他细胞为食，其营养供给来自于祖先遗留下来的叶绿体。相应地，植物细胞虽然保持了可以负责移动的细胞骨架，但失去了迅速改变细胞形态和

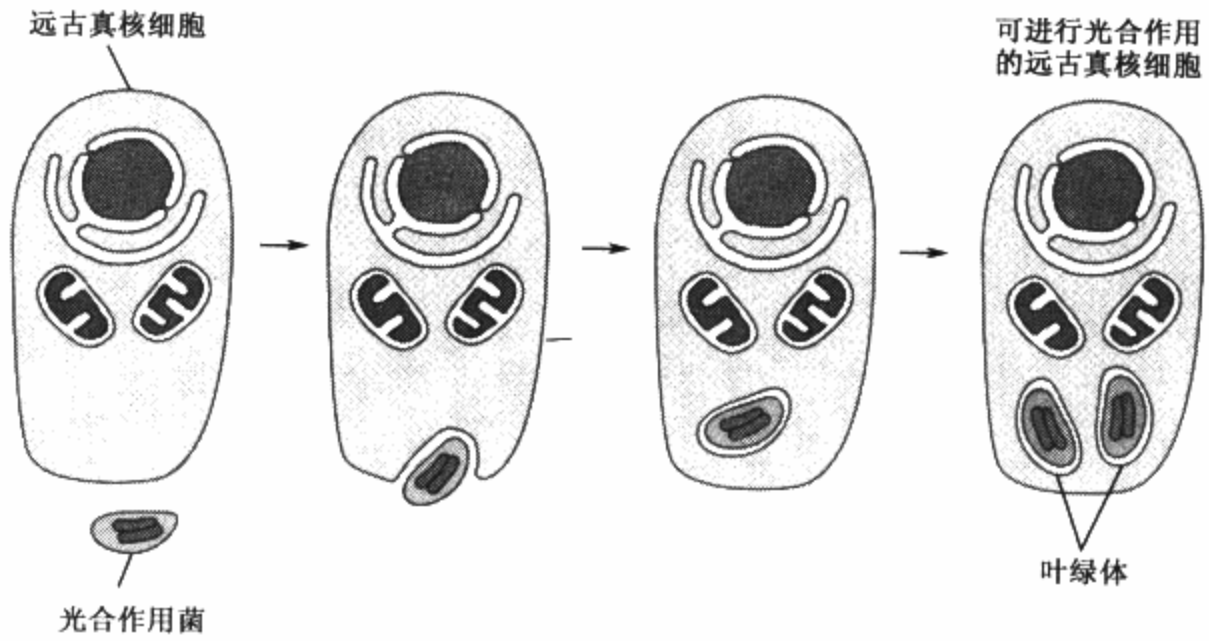


图 1-37　叶绿体的起源

已经含有线粒体的远古真核细胞，吞噬了一种光合作用菌，形成共栖系统。目前人们认为叶绿体依旧含有 10 亿年前其祖先光合作用菌的内共生体特征

吞噬其他细胞的功能。并且它们围绕自身产生了一层坚硬起保护作用的细胞壁。如果远古的真核生物果真是以其他物种为食的捕猎者，那么我们可以认为植物细胞就是从狩猎者到农业者过渡的真核细胞。

真菌代表了另外一种真核生物的生存方式。真菌细胞和动物细胞一样含有线粒体但没有叶绿体；但是和动物细胞及原生动物不同的是它们有一层坚硬的外壁使它们不能迅速移动吞噬其他细胞。真菌好像是由猎人到清洁工的转变：它们以其他的细胞分泌或者通过死亡来释放的营养分子为食，它们所有的消化都是通过往体外分泌消化酶来完成的。

真核生物有着杂和的基因组

真核生物的基因组原本就是一个杂合的——有来自于远古厌氧真核细胞的部分，也有共生的细菌部分的。大多数的遗传信息储存于细胞核中，但是有一小部分存于线粒体中，对于植物和藻细胞来说还有一部分存于叶绿体中。线粒体和叶绿体 DNA 可以单独地和细胞核 DNA 分离开进行分析和测序。研究者发现线粒体和叶绿体基因组相对于细菌基因组来说是一个退化的、消减的版本，其中缺少许多具有重要功能的基因。例如，在人类细胞中，线粒体基因组仅含有 16 569 个碱基，编码 13 种蛋白质，仅有两种核糖体 RNA 和 22 种转运 RNA。

线粒体和叶绿体中丢失的基因并没有完全丢掉，其中许多由于未知的原因由共生细菌中转移到了宿主细胞核中。人类细胞的细胞核 DNA 中含有许多线粒体中重要蛋白质的编码基因；在植物中，细胞核 DNA 中也含有许多叶绿体所需蛋白质的编码基因。

真核生物的基因组非常庞大

自然选择赋予了线粒体小型的基因组，就像细菌有着小型基因组一样。相反地，真核生物的细胞核基因组好像自由扩充了许多。也许真核生物的生存方式使得这种庞大的规格更有优势：捕食者需要比猎物更大的体积，以及更大的基因组。无论什么样的解释，大多数真核细胞比细菌和古细菌大是不争的事实（图 1-38）。含有极大量的 DNA 也应有其更深的含义。

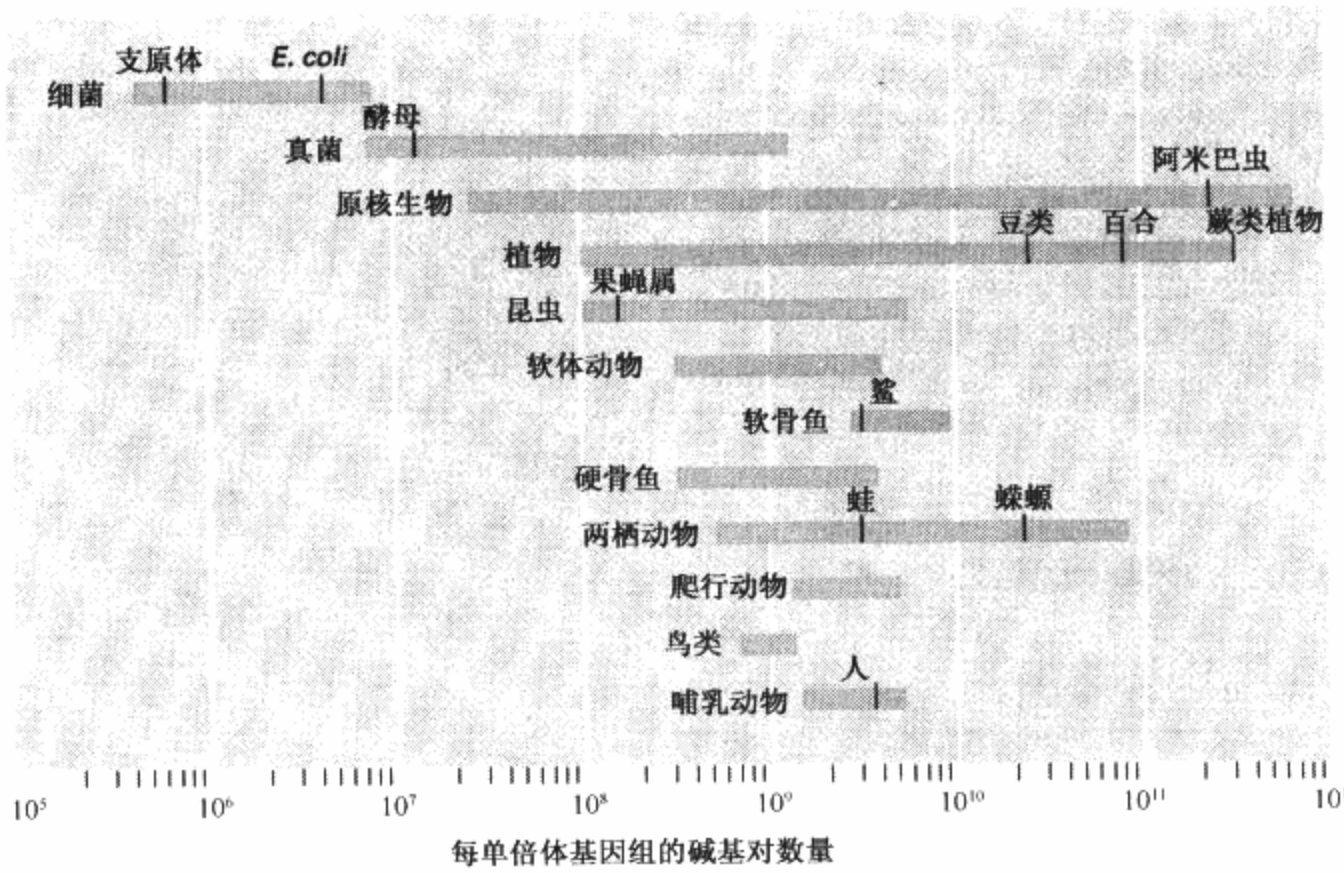


图 1-38　基因组大小比较

基因组大小依据单倍体基因组的 DNA 核苷酸数量来计算，即一个基因组的拷贝。（进行有性生殖的物种，像我们自己的细胞通常都是双倍体：含有两套基因组的拷贝，一套来自于父亲，另一套来自于母亲。）关系较近的物种在基因组 DNA 的量上却差别很大，即使它们含有相似数量的功能基因（数据摘自 W.-H. Li, Molecular Evolution, 380～383. Sunderland, MA: Sinauer, 1997.）

真核生物不仅基因数远大于原核生物，并且有着更大量的不编码蛋白质或其他功能分子的 DNA。人类基因组的碱基数是典型细菌的 1000 倍，基因数目是 20 倍，非编码 DNA 数目是 10 000 倍（人类基因组中大约 98.5%的 DNA 是非编码的，而 *E. coli* 中仅有 11%）。

真核基因组中有着丰富的调控 DNA

在非编码 DNA 中，许多都几乎是不必要的，就像一大批过期的刊物，其出现是因为几乎没有保持小型信息记录的压力，所以保留所有的信息就比分出有价值信息而遗弃其他无用信息要容易得多。也有些例外的真核生物，如河豚（图 1-39），不像其他物种有着这种遗传信息的浪费，它们不仅可以清除自身的非编码 DNA，还可以避免在第一时间内获得这种遗传信息。然而它们却和亲缘关系比较近但有着相对非常复杂 DNA 的其他物种有着相似的外形和行为。

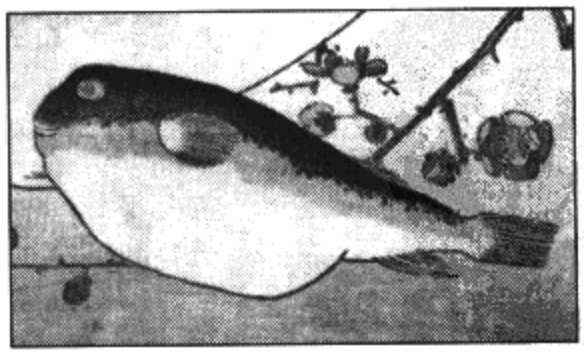


图 1-39　河豚

河豚的基因组仅有 4 亿个碱基——大约仅有斑马鱼的 1/4，然而两个物种含有相近数量的基因（摘自 a woodcut by Hiroshige，经 Arts and Designs of Japan 许可）

即使在像河豚这种基因组压缩的真核生物中，非编码 DNA 的数量也远远多于编码 DNA 的数量，但至少一些非编码 DNA 有着重要的功能。它们有的可以调控相邻基因的表达。有了这种调控 DNA，真核生物进化出了特殊的方式来调控某种基因何时以及何地发挥作用。这种复杂基因调控模式对于复杂多细胞机体的形成有着重要作用。

基因组决定多细胞发育的进程

单个动物或者植物体内的细胞是多种多样的。脂细胞、皮细胞、骨细胞以及神经细胞，它们千差万别。然而所有类型的细胞都是在胚胎发育时由单个受精卵细胞产生的，所有的细胞（极少数例外）都含有本物种基因组的相同拷贝。

一种解释认为，细胞可以根据从环境中接受到的信息有选择地利用自身的遗传指令。DNA 不仅仅是一个决定每种细胞所必须含有的分子购物表，细胞也不是所有购物表上东西的集合。正确地说，细胞应该是一个有着多重指令的机器，可以感知环境信号，并且有着很强的能力，可以根据接受到的信号唤醒一系列的基因发挥作用。许多真核生物的基因可以编码调控其他基因表达的蛋白质。大多数基因调控蛋白质通过直接地或间接地结合与被调控基因相连的调控 DNA 序列而发挥作用（图 1-40），或者通过干扰其他蛋白质来完成这项工作。真核生物扩充的基因组不仅指定了细胞内的硬件设施，而且还安装了可以控制硬件怎样发挥作用的软件设施。

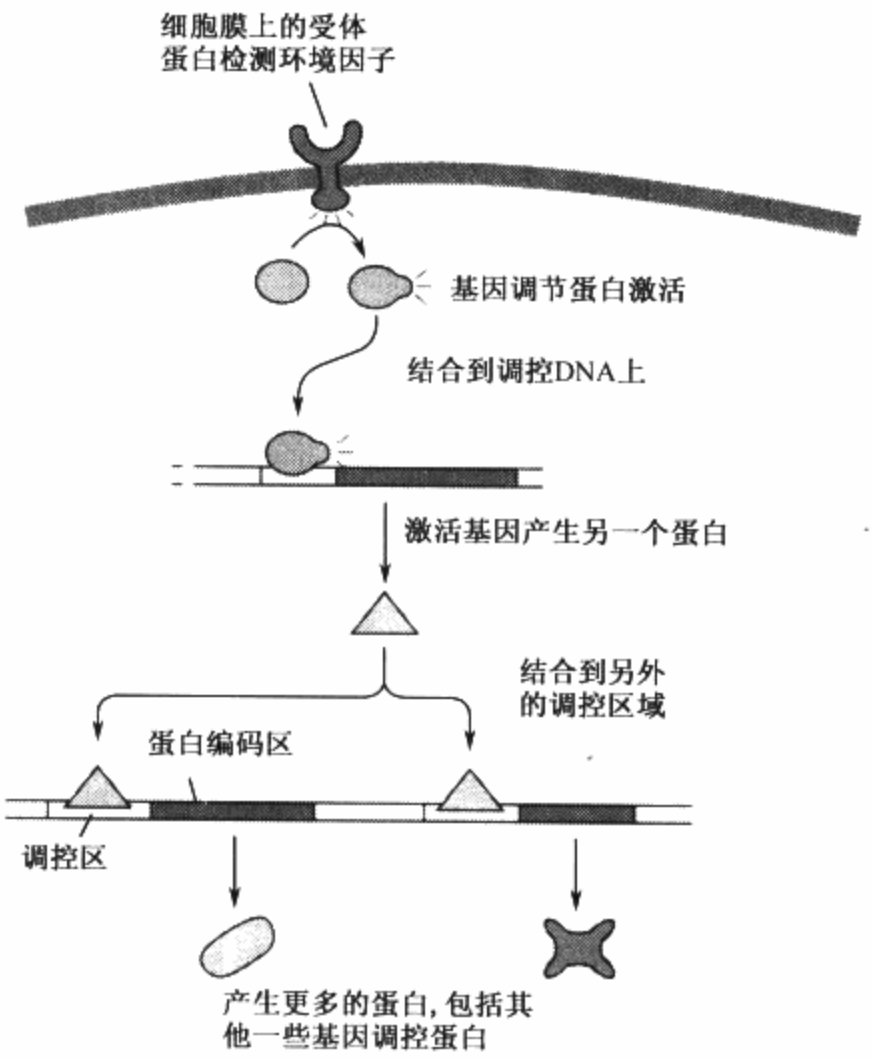


图 1-40　环境信号调控基因的表达

调控 DNA 使得基因的表达可受调控蛋白质的控制，这又可依次调控其他基因的表达。这张图显示了细胞内基因的表达是怎样受环境信号的调节的。起初外界信号只激活细胞内调控蛋白质，如信号可触发调控蛋白质磷酸基团并改变其化学特性

在发育中的多细胞机体中，每个细胞都由相同的控制系统所管理，但是其结果依赖于细胞和其临近细胞进行交换的信号。结果是令人惊异的，可以产生不同状态细胞的精细图案，每种细胞表现出跟其在多细胞机体中所处位置相应的特征。基因组足够庞大，可以调节控制整个多细胞有机体的遗传信息，但在单个细胞中只利用了其中一部分信息（图 1-41）。

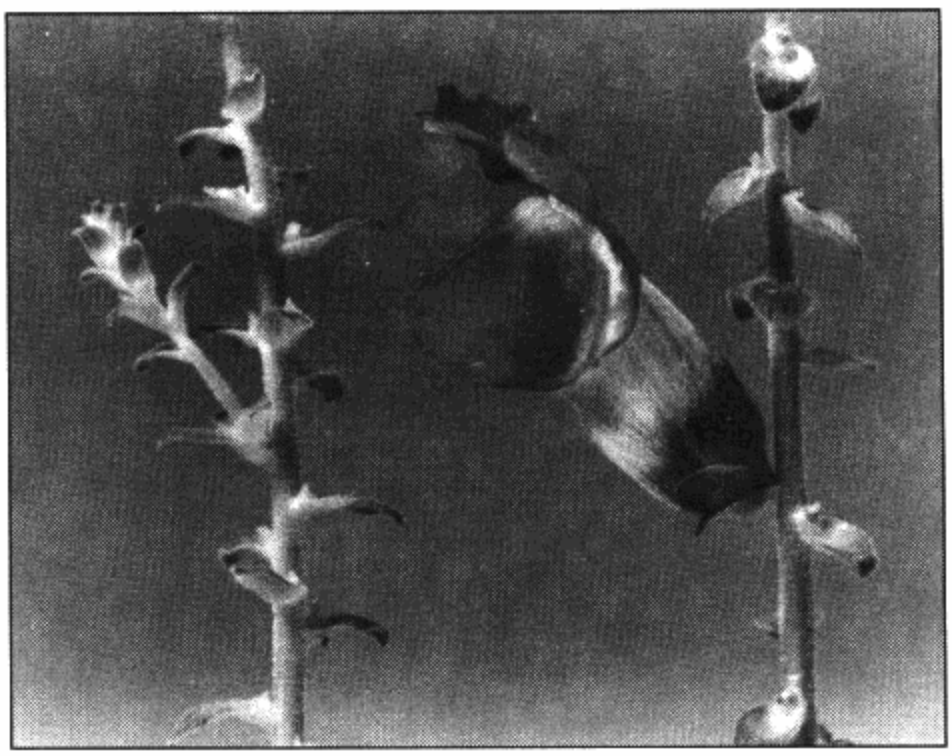


图 1-41　多细胞发育的遗传调控

在金鱼草（*Antirrhinum*）中证实了调控基因的功能。此图中，一个编码调控蛋白质的基因突变使得原本开花的位置生出了一个叶枝：因为调控蛋白质的改变，细胞选择了和正常植物不同的发育位置。左边是突变体，右边是正常植株（经 Enrico Coen 和 Rosemary Carpenter 许可）

许多真核细胞以单细胞的形式存在，即原生生物

多数真核细胞以单细胞的形式存在，一些作为捕猎者（原生动物），另一些是光合作用者（单细胞藻类），还有一些作为清洁工（单细胞真菌或酵母）。即使不形成多细胞结构，单细胞真核生物，或者说原生生物，也是非常复杂的。图 1-42 显示了它们所具有的多种外形。特别是原生动物的解剖学表明，它们的结构非常精细，其中包括感觉鬃、光感应器、纤细的纤毛、腿状附属物、口腔、螯刺以及肌肉状的伸缩性维管束。尽管原生动物是单细胞的，但它们和许多多细胞生物一样有着复杂而精细的行为（图 1-33）。

从其祖先以及 DNA 序列上来看，原生生物和多细胞真核生物的三大分支——动物、植物和真菌有着极大的不同（图 1-21）。和原核生物一样，人们容易忽视原生生物，因为它们是肉眼所不能见的。直到现在，人们依靠基因组学的帮助，才开始了解它们在生命树中的位置，了解它们在我们漫长的进化中所起到的作用。

酵母是最小的真核模式生物

真核生物分子以及遗传上的复杂性是令人震惊的。即使生物学家们在真核生物中投入的精力远大于原核生物，但也只能把有限的资源用于研究几种选定的模式生物。

在研究真核细胞内在的活动时，为了排除多细胞发育所造成的影响，使用尽可能简单的单细胞种类显得合理得多。生物学家们选择了最简单的模式生物——酿酒酵母（图 1-43），啤酒酿造者和面包师选择的种类来进行研究。

酿酒酵母是真菌世界中一类小的单细胞生物。现在认为，它和动物、植物的亲缘关系都非常接近。酵母十分强壮，可以在简单培养基上生长。和其他真菌一样，它有着坚硬的细胞

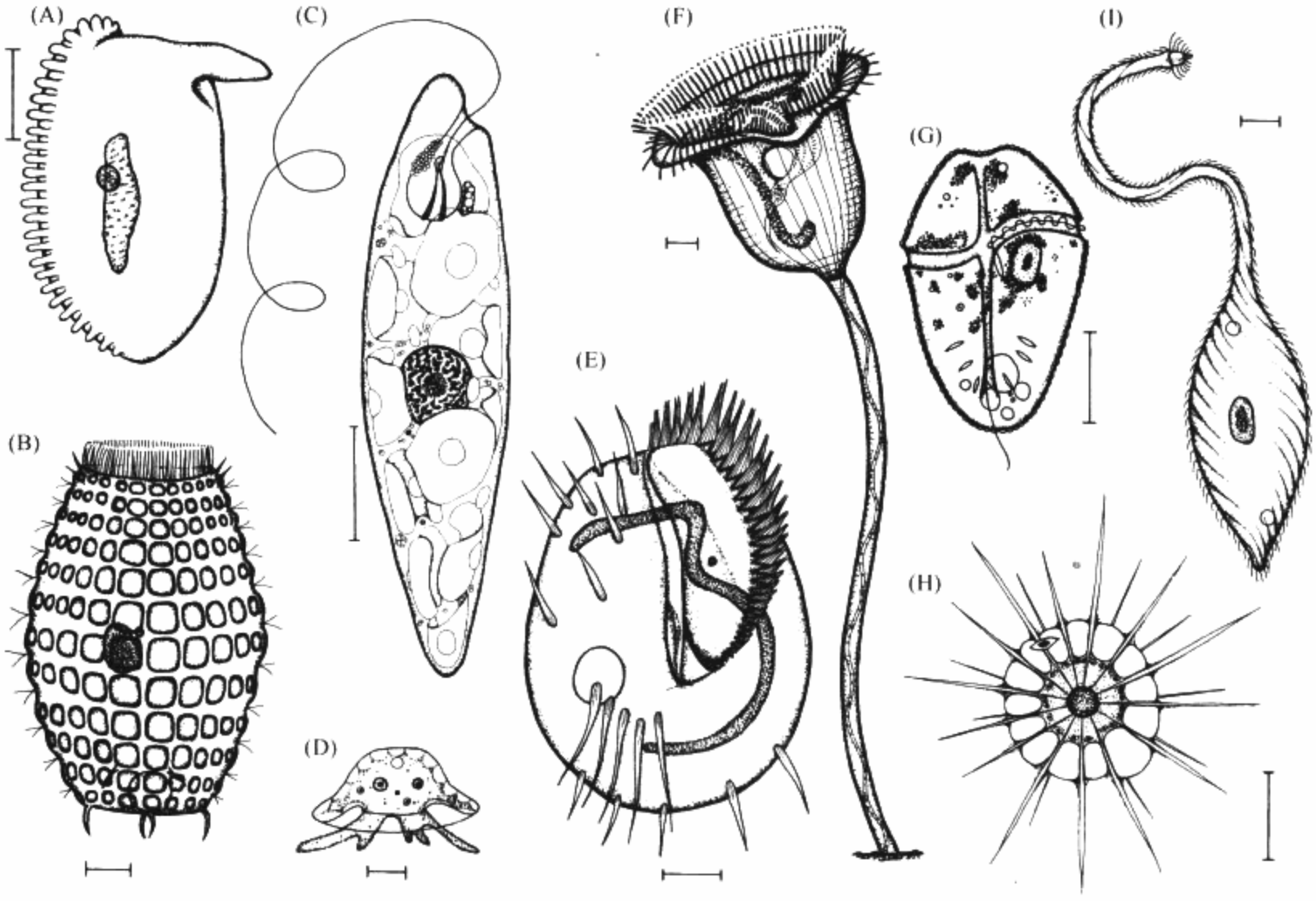


图 1-42　原生动物的分类：一类极端分化的物种的几个小个体

简图按照不同的尺度勾画，每一个图中刻度尺都代表 10μm。（A）、（B）、（E）、（F）和（I）是纤毛虫；（C）是眼虫；（D）是变形虫；（G）是腰鞭毛虫；（H）是太阳虫（摘自 M. A. Sleigh，Biology of Protozoa. Cambridge，UK：Cambridge University Press，1973.）

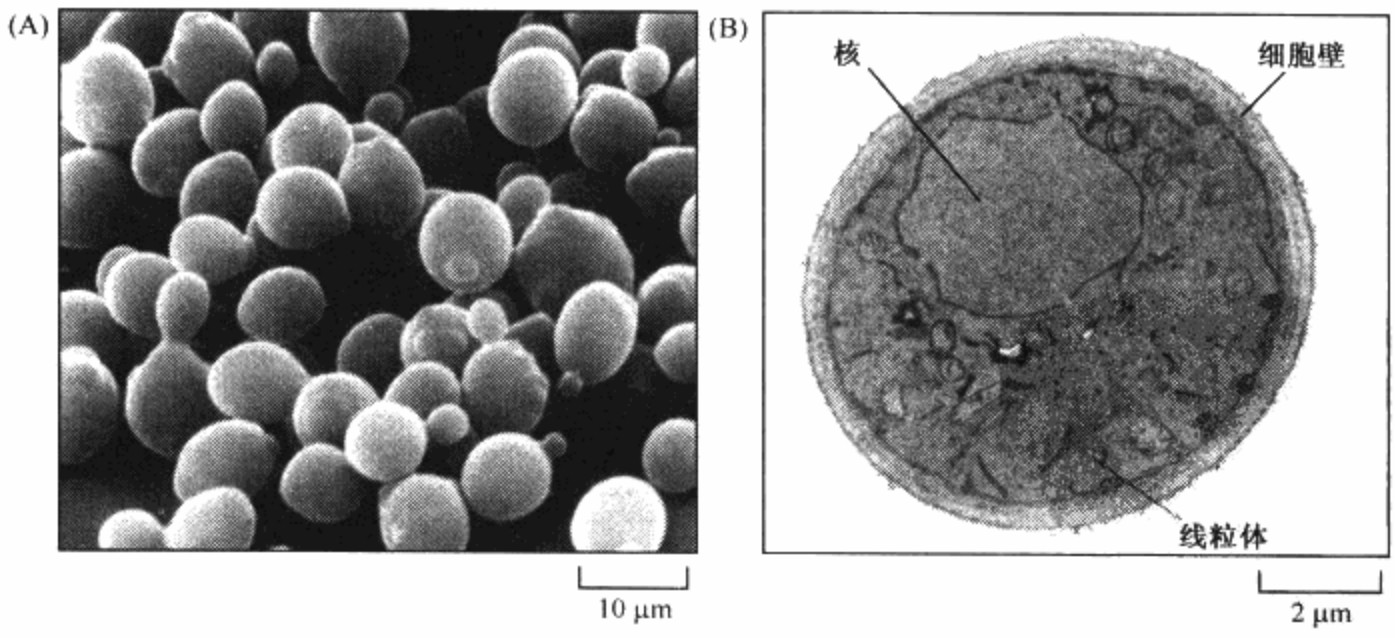


图 1-43　酵母（*Saccharomyces cerevisiae*）

（A）一簇细胞的扫描电镜照片。这是一种出芽酵母；通过从原细胞发出小芽并分离出来而完成生殖。电镜能见许多出芽的细胞。（B）酵母细胞发射电镜下的横切面，显示出了细胞核、线粒体以及薄的细胞壁［（A）经 Ira Herskowitz 和 Eric Schabatach 许可］

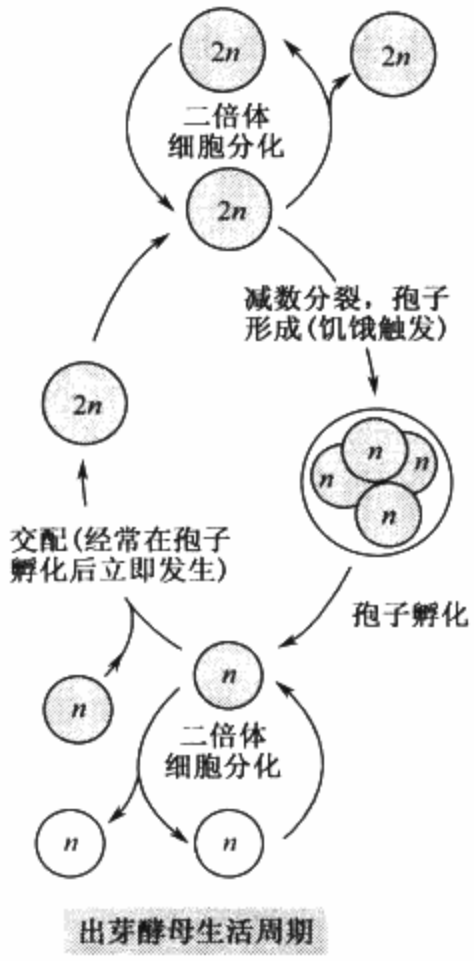


图 1-44　酵母的生活周期

根据环境条件以及基因型的不同，酵母存在两种生活阶段：一种是含有两套染色体组的二倍体状态，另一种是仅有单个染色体组的单倍体状态。二倍体的形式可以进行正常的细胞分裂或者减数分裂来产生单倍体细胞。单倍体形式的细胞可以进行正常的细胞分裂周期或者选择和另外一个单倍体细胞进行有性融合形成二倍体细胞。在食物缺少的情况下，减数分裂被启动，并且形成孢子——一种休眠状态的单倍体细胞，对恶劣的环境条件有抵御力

壁，相对比较稳定，含有线粒体但没有叶绿体。当营养丰富时，它们可以像细菌一样快速生长、分裂。它们可以进行无性（即通过简单的细胞分裂）或有性生殖：两个单倍体的酵母细胞（含有基因组的单个拷贝）融合在一起产生一个双倍体细胞（含有一对基因组）；双倍体细胞又进行减数分裂（一种减少的分裂）产生单倍体细胞（图 1-44）。和高级的动物、植物相反，酵母在单倍体还是多倍体阶段进行分裂是不确定的，它们可以随意地根据生长环境来进行两种状态的互换。

酵母还有其他的特征如其基因组是真核的，但却出乎意料的简单，所有的这些特征都使得酵母作为遗传学研究上的一种非常方便的物种。虽然简单，但是它们必须完成所有真核细胞必须执行的任务。我们以后将会讲到，酵母的研究（酿酒酵母和其他种类）为了解许多重要的过程提供了一把钥匙。其中包括真核生物的细胞分裂周期——细胞生活中的重要事件，细胞核和其他细胞组分进行复制并且分配到两个子细胞中去。控制这个事件的系统在进化过程中是非常保守的，以致在酵母和人类细胞中执行这种功能的许多组分是可以互换的：如果缺少了某种细胞分裂周期必需蛋白质组分的酵母中转入了人类细胞分裂周期蛋白质的同源物，酵母就会恢复正常的分裂。

机体中所有基因的表达水平都同时受到监控

酿酒酵母的全基因组序列于 1997 年测出。其中含有 13 117 000 个碱基，包括线粒体的 78 520 个碱基。碱基数目仅为大肠杆菌的 2.5 倍，蛋白质数目仅为大肠杆菌的 1.5 倍（总计有 6300 个）。酿酒酵母的许多生活方式都和细菌的一样，说明酵母也可能经历了使其基因组紧缩的选择压力。

某些物种全基因组序列的信息，如酵母和人类，为细胞工作开阔了可观的前景：以前感觉不可能的事情现在则尽在我们的掌握之中。举个例子来说，现在可以把酵母体内编码 6300 个蛋白质的相应基因点在一个小的玻璃片阵列中。由于 DNA 分子可以通过碱基配对结合到其互补序列的 DNA 或 RNA 分子上，所以如果核酸分子的混合物通过这样一个 DNA 阵列时，混合物中的每种分子都可以结合到相应的 DNA 点上。如果混合物中的分子都连有一个荧光标记的化学标签时，那么我们就会得到一系列的荧光斑点，分别对应于阵列上的特定点（图 1-45）。按照这种方式，我们就可以同时监测某个特定时期酵母中的所有基因的 mRNA，并且可以观察到当环境变化时所有基因的变化。这种方法还可以用于分析缺失了某种基因的突变体中我们感兴趣的任何基因。原则上来说，这种方法可以揭示出整个控制基因表达的系统，不仅酵母细胞中，而且可以用于任何一种已得到基因全序列的物种。

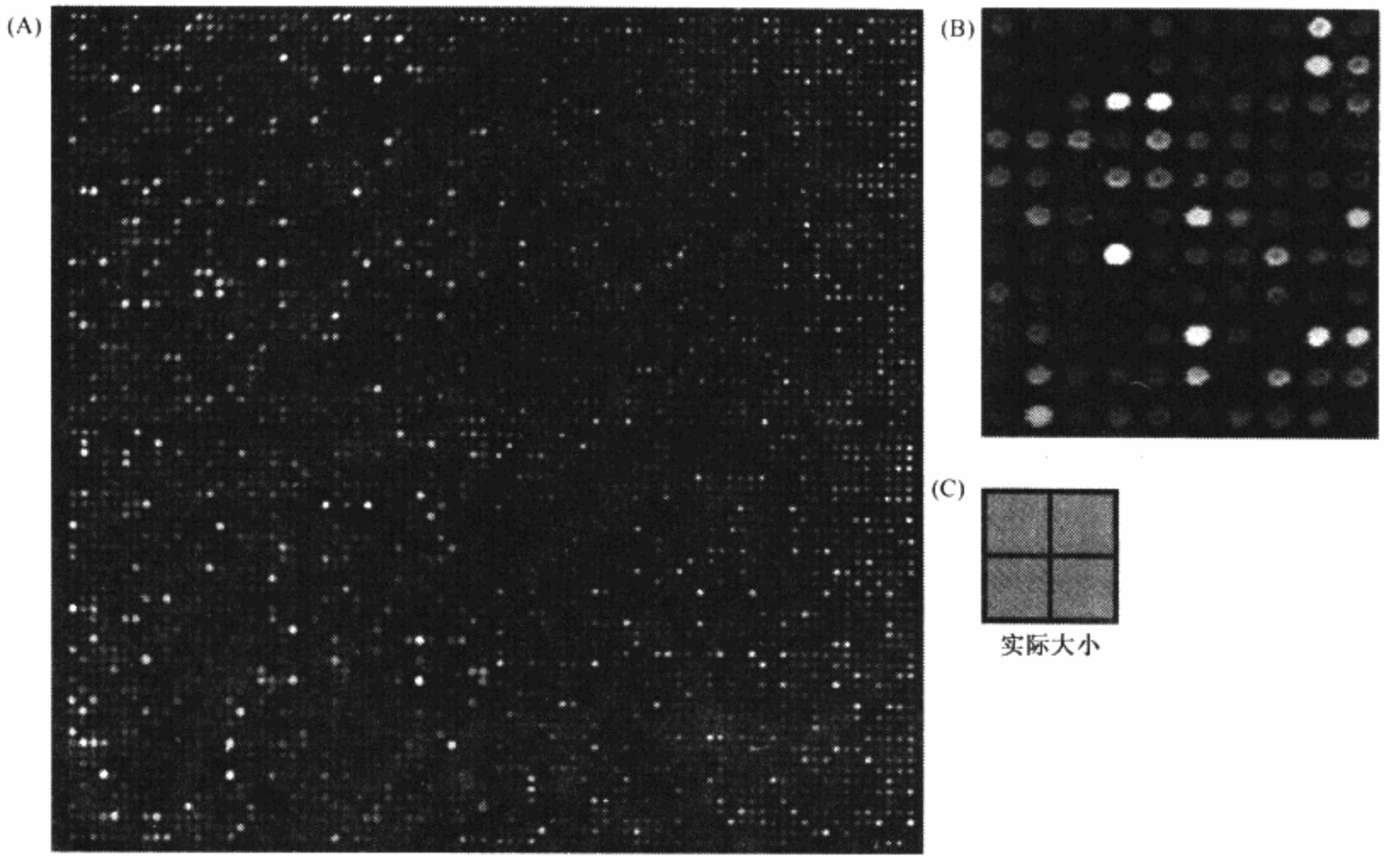


图 1-45　利用 DNA 阵列研究酵母基因表达的变化

酵母基因组中所有 6300 个编码蛋白质的基因 DNA 样品储存在一个玻璃载玻片的方形阵列中：（A）显示了含有 6400 个点的阵列；（B）显示了高度放大后的一小部分区域；（C）显示了阵列（A）的实际大小。这张阵列是把培养基中有和无葡萄糖的酵母的一系列 mRNA 分子进行比较。如果含有葡萄糖培养的细胞转录了特殊的 mRNA，那么细胞中的 mRNA 样品中就含有可以结合到阵列中相应 DNA 点的分子。实际操作上，提取出的 mRNA 分子并没有直接用于阵列的杂交上（会在第 8 章解释）而是生成了另外与其对应的一种叫做 cDNA 的核酸分子。这个试验中，cDNA 利用绿色荧光来标记。当 cDNA 混合物和阵列杂交时，每个 cDNA 分子都结合到阵列上相应的点上：某个点上强的绿色荧光代表对应的基因有一个强表达。接着对不含葡萄糖培养的酵母细胞进行以上操作，把其 cDNA 利用红色荧光标记，然后用于同一张阵列。因此，绿色的点代表在葡萄糖存在时大量表达的基因；红色的与其相反。在两种情况下基因表达相同的基因可以同时结合绿色和红色荧光的 cDNA，因此呈现黄色（摘自 J. L. DeRisi et al.，*Science* 278：680～686，1997. © AAAS.）

拟南芥——300 000 多种植物的模式物种

我们身边的大型多细胞物种（花草树木以及动物）好像千差万别，但它们相对于单细胞微生物来说，在进化的起源上更为接近，并且基础细胞生物学也更相似。30 亿年前细菌和真核生物就发生了趋异进化，而脊椎动物和昆虫 7 亿年前才产生分化，鱼类和哺乳动物 4.5 亿年前分化，不同种类的显花植物仅在 1.5 亿年前才开始分化。

因为所有的显花植物都有着相对较近的进化关系，我们可以再次选择一种或者几种植物进行详细分析进而探索这一大类物种的细胞和分子生物学。分子生物学家们从几十万种显花植物中选出了一种小型杂草——十字花科植物拟南芥（图 1-46）。拟南芥可以大批量的种植在室内，每 8～10 个星期可以产生上千株后代。其基因组大约有 1.4 亿个碱基，是酵母的 11 倍，目前其全基因序列已经测出。

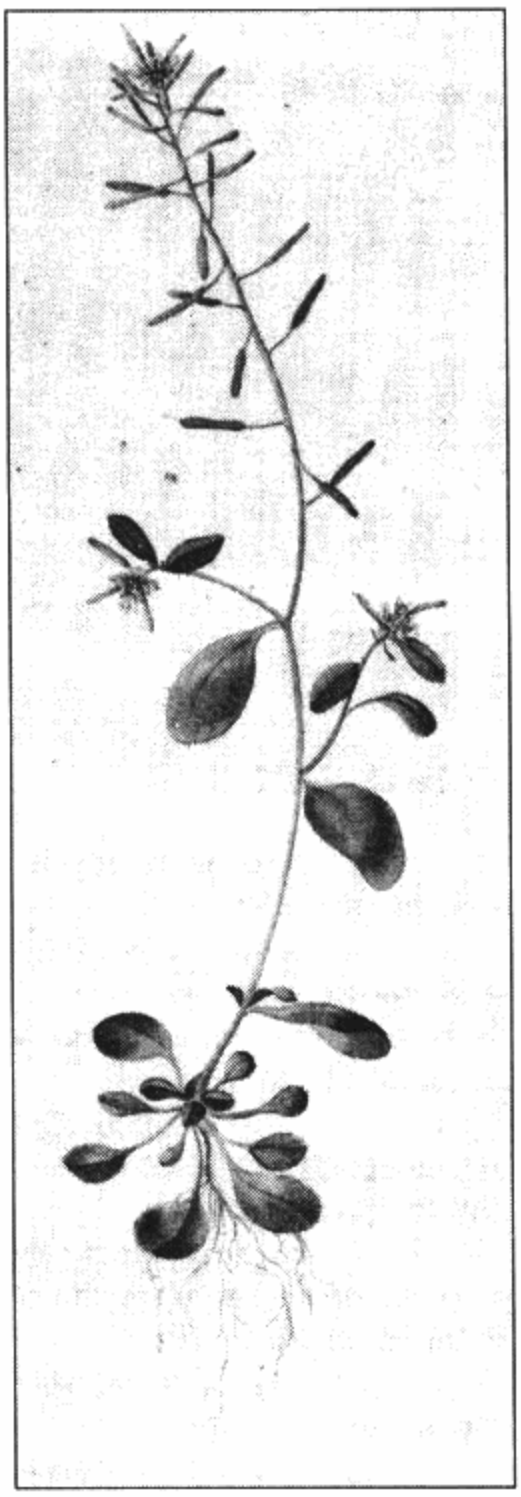


图 1-46　拟南芥——研究植物分子遗传学的基本模式植物

（经 Toni Hayden 和 the John Innes Foundation 许可）

动物细胞的代表物种：线虫、果蝇、小鼠、人

多细胞动物占据了已经命名的大多数生物，并且是大多数生物学研究的对象。人们选择 4 个物种作为最初的分子遗传学模式生物，目前都已得到它们的全序列。它们是秀丽新小杆线虫（*Caenorhabditis elegan*）、黑腹果蝇（*Drosophila melanogaster*）、小鼠（*Mus musculus*）和人类（*Homo sapiens*）。

C. elegan（图 1-47）是一种小型的、损害庄稼的线虫的无害近缘种。由于其生活周期仅有几天，适合于遗传学研究，并且可以在冰箱中以生存暂停状态无限期的存活。线虫有着简单的身体结构，是一种理想的模式生物，并且有着机械式的精确发育，可以由一个受精卵发育成一个准确含有 959 个体细胞（还有一定数量的卵细胞和精子细胞）的成虫——所有的动物中都没有这种不同寻常的一致性。目前我们已经根据严格的方式掌握了线虫细胞分裂、移动以及改变自身形态的详细数据。线虫的基因组有 9700 万个碱基，编码 19 000 个蛋白质，并且目前已经有了多种突变体的资源，可以用于研究特定基因的功能。虽然线虫的身体结构和我们的不同，但其体内许多保守的生物学机制却为我们提供了了解人体多种进程的宝贵模式。例如，线虫的研究使我们了解决定体内细胞数量的细胞分裂和细胞死亡的过程——这是发育生物学和癌症研究的重要课题。

果蝇的研究为脊椎动物发育学提供了钥匙

果蝇（图 1-48）作为模式生物研究的时间远远长于其他物种，实际上，大部分经典遗传学都是建立在此种昆虫的研究之上的。例如，在 80 多年前，果蝇研究者就已经通过光学显微镜为基因（遗传信息的主要单位）存在于染色体上提供了充足的证据，染色体在真核细胞中的行为是相近的，但其功能一开始还未知。证据来源于果蝇的一项模式生物特征——巨染色体，该染色体有着特异的花纹外形，在一些果蝇细胞中可见（图 1-49）。人们发现突变果蝇家族中遗传信息的特殊变化，是和巨染色体上某个花纹的改变直接相关的。

近年来，相对于其他生物，果蝇更多的为我们提供了怎样追踪染色体 DNA 中的遗传指令和成熟多细胞机体间的因果关系。身体器官错位的果蝇突变体使我们可以发现及鉴定控制身体器官，如肠道、眼睛、四肢或其他部分在身体上正确定位的基因。如果参与上述调控的基因得到测序，就可以对脊椎动物的基因组进行比对寻找同源基因。找到同源基因后，我们就可以通过分析小鼠中的突变体进行基因功能的验证。最终的结果揭示昆虫和脊椎动物的发育的分子机制有着令人诧异的相似性，这一点我们将在本书的后续章节中讲到。

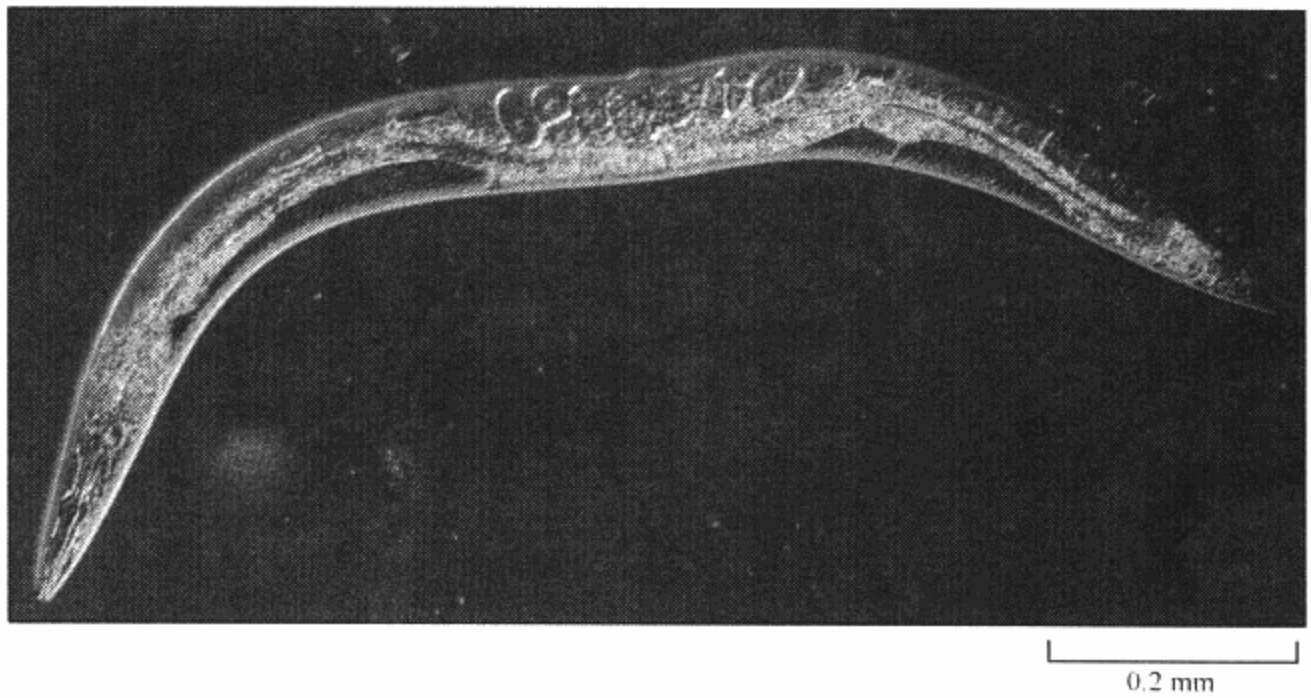


图 1-47　秀丽新小杆线虫是第一个测定全基因序列的多细胞生物

这种小的线虫，只有 1mm 长，生活在泥土中。大部分个体是双性的，可以同时产生精子和卵子（经 Ian Hope 许可）

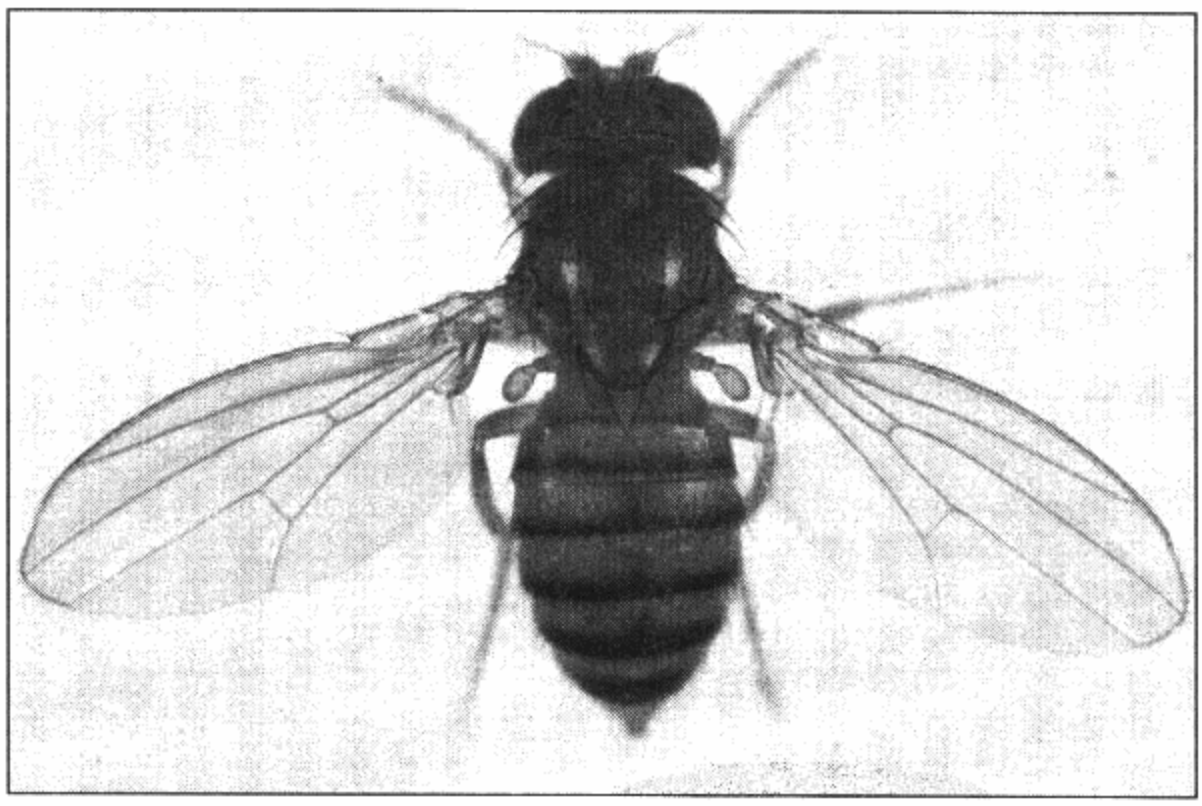


图 1-48　黑腹果蝇

对应这种果蝇的研究为人们了解所有的动物是怎样由一个受精卵发育为成熟个体提供了有力工具（摘自 E. B. Lewis, *Science* 221：cover，1983. © AAAS.）

已命名的大部分物种都是昆虫。所以即使果蝇仅仅和昆虫有相似性而和脊椎动物差异较大，它也是一个重要的模式生物。然而既然了解脊椎动物的分子遗传学是根本目的，为什么不进行一步到位的研究呢？为什么要通过研究果蝇进行间接的了解呢？

果蝇由一个受精卵发育成成体只需 9 天时间；饲养果蝇要比任何一个脊椎动物要容易得多，并且基因组要小得多，只有 1.7 亿个碱基，而人基因组有 32 亿个碱基。果蝇编码 14 000 个蛋白质，有着许多基因的突变体。还有另外一个更深层的原因，即在脊椎动物中难以揭示

的遗传机制在果蝇中比较容易发现。这种原因和基因复制的频率有关，即脊椎动物基因组中基因复制频率要远远大于果蝇，这就决定了脊椎动物物种的复杂性和精密性。

脊椎动物基因组是重复复制的产物

几乎每个真核生物基因组中的基因都有同源物存在，即同一基因组中明显的相似基因，它的出现一定是基因复制的产物。许多情况下，几种基因在整个基因组中的比对结果相似，表明基因是以连续的群体而不是一个单独的个体进行复制的。有一种假设来解释这种现象，在脊椎动物进化初期，全基因组接连进行了两次复制，这样每个基因就有了 4 个拷贝。在一些脊椎动物中，例如，鲑鱼和鲤鱼家族（其中还包括斑马鱼，一种目前研究比较热的模式生物），预计多出了一轮复制，每个基因产生了 8 个拷贝。

由于长期进化过程中更进一步的进化变迁，脊椎动物基因组进化的具体过程还不是很清楚。起始相同的基因存在着分化，基因的许多拷贝经过分裂性的突变而丢失了；一些拷贝继续着开始的复制；每一个脊椎动物家族的分支物种的基因组都重复着基因重组，最终打破了远古的最初基因模式。两个相近物种间的仔细比对，如人和小鼠，显示出在脊椎动物的进化

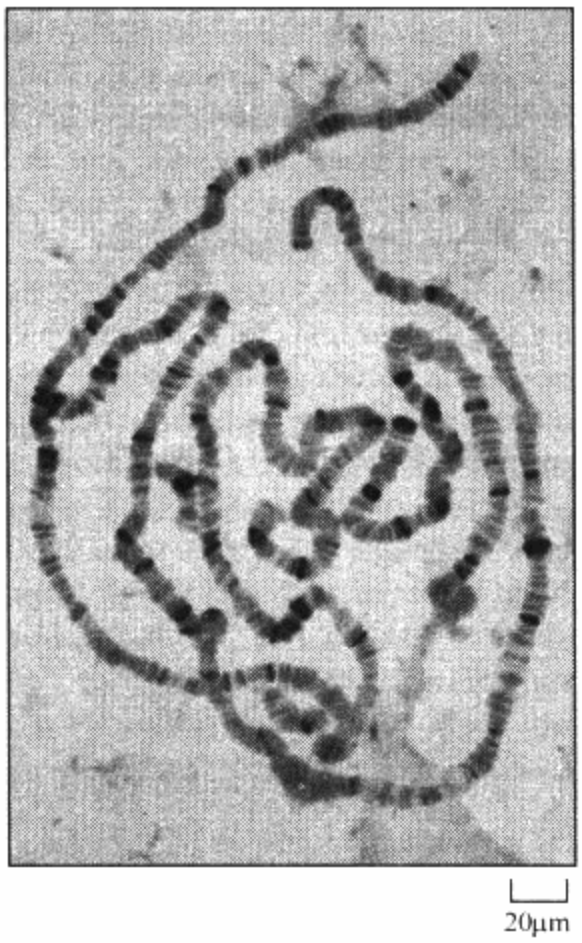


图 1-49　果蝇唾腺细胞的巨染色体

由于 DNA 在细胞没有分裂的情况下进行了多轮复制，所以每个细胞染色体中都含有 1000 个相同的 DNA 分子，并且都对齐排列。这使得它们在光学显微镜下清晰可见，呈现出清晰的明暗相间的花纹。特定的花纹代表特定基因的位置：丢失了一个花纹的突变果蝇显示出丢失相应基因的表型。转录速度很高的基因其位置呈现一种泡状形式。显微镜下暗褐色的条带是特殊调控蛋白质结合 DNA 的位置（摘自 B. Zink and R. Paro, from R. Paro, *Trends Genet*. 6: 416～421, 1990. © Elsevier.）

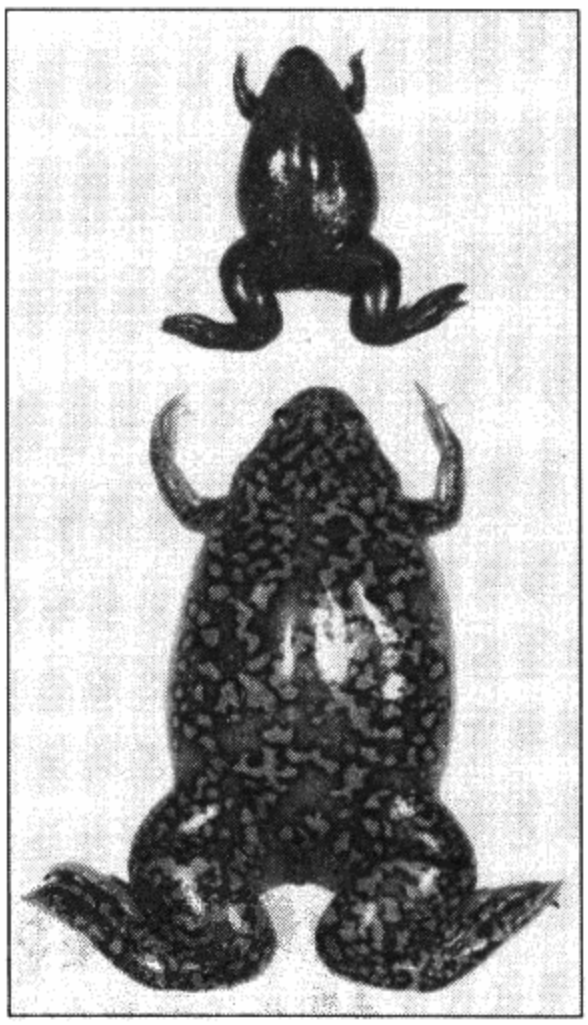


图 1-50　蛙（*Xenopus*）的两种种类

上面的 *X. tropicalis* 有着正常的二倍体基因组；下面的 *X. laevis* 每个细胞有着成倍的 DNA。通过染色体的带型、基因的排列以及基因序列的对比，人们清楚地了解到高倍体的物种是由全基因组的重复而进化得来。这种重复是由于分化较小蛙的不同种间杂交产生的（摘自 E. Amaya, M. Offield, and R. Grainger, *Trends Gen*. 14: 253～255, 1998. ©Elsevier.）

史上，染色体经常融合或断裂来移动 DNA 的大片段。实际上，就像我们将在第 7 章中讲到的那样，目前基因组的格局完全有可能是许多基因组片段独立复制的产物而不是整个基因组统一复制形成的。

然而可以确定的是，在进化过程中整个基因组统一复制的情况也是随时发生的，我们还可以经常清楚地看到这种基因组复制单元。例如，蛙属（*Xenopus*）就由多种含有 2 倍或者 3 倍基因组的种组成。其中 *X. tropicalis* 是正常的双倍体基因组，实验室种 *X. laevis* 每个细胞有重复的基因组，DNA 是正常细胞的 2 倍，*X. ruwenzoriensis* 是 6 倍体细胞（有 108 条染色体，而 *X. laevis* 有 36 条染色体）。这些种类估计都是在 1.2 亿年前起源于同一个种（图 1-50）。

基因冗余对于遗传学家来说是个大问题，但却为进化中的物种提供了机会

虽然不清楚进化的具体细节，但是人们却清楚地看到，在脊椎动物的基因组中，大多数原来相同的基因目前却以不同的形式存在。相关的基因可以为了某些目的进行功能上的互换。这种现象叫做**基因冗余**。这对于那些奋力研究基因进化的特殊过程的科学家来说是个头疼的问题。如果基因 A 突变后没有产生表型，我们也不能说它是功能不相关的——有种可能是此基因和其相关基因有着相同的功能，这样即使基因 A 突变后也不会失去正常的功能（就像一架坏了其中一台引擎的客机照样飞行一样）。果蝇重复基因组相对较少，基因重复也较少，所以分析起来就显得比较顺利：单个基因功能可以通过此基因的突变直接看到（就像单引擎的飞机引擎坏掉之后就不能再飞行了一样）。

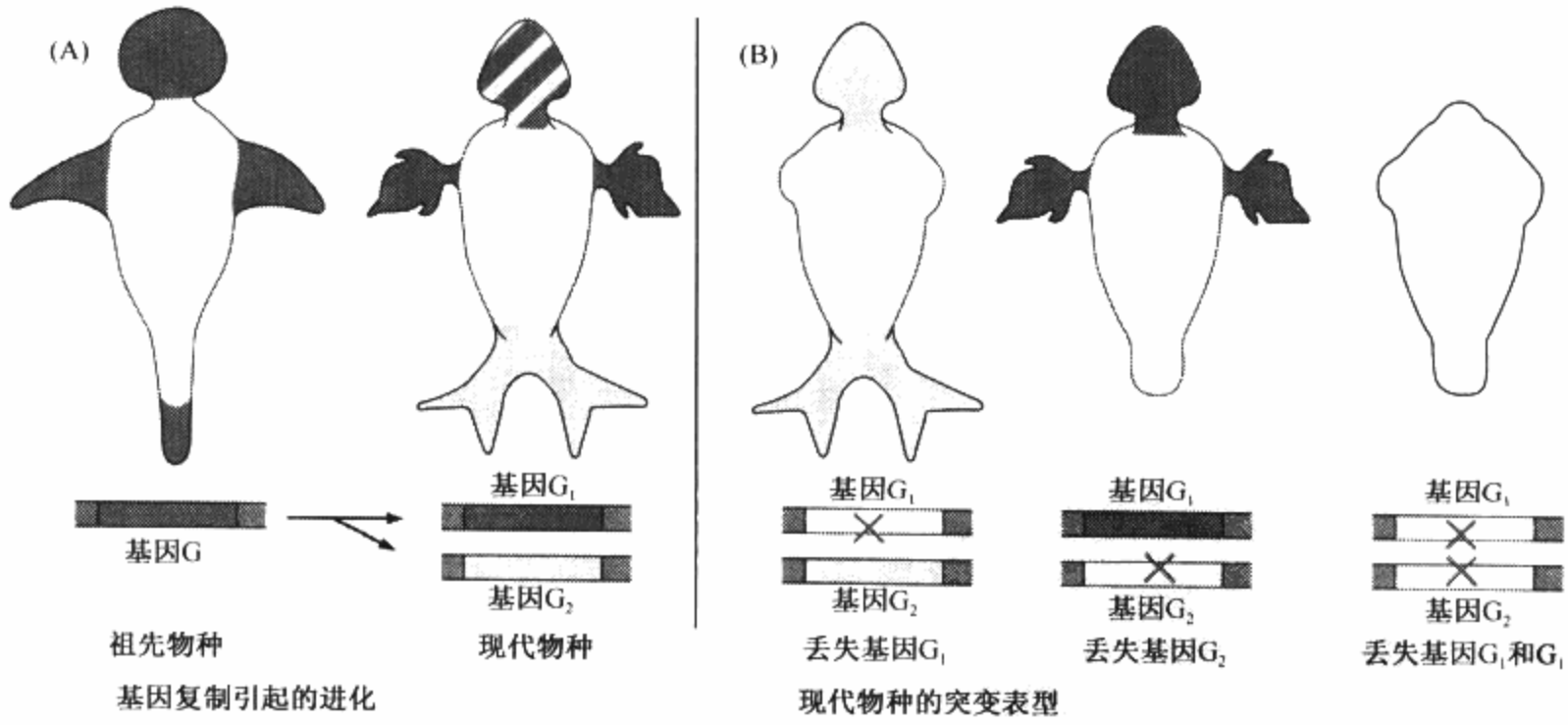


图 1-51　基因复制对于研究基因功能的突变分析的后果

在这个假设的例子中，祖先多细胞物种有基因 G 的单拷贝，此基因在身体的多个位置发挥功能，以绿色标出。（A）经过基因重复，子代生物有了基因 G 的两个拷贝，G_1 和 G_2。它们在其作用位置的表达多少有些不同，但依旧保持了重要作用的相同。在一些位置，它们共同表达，每个拷贝独立地行使其祖先基因 G 所执行的功能（交替的绿色和黄色条）；在另一些位置，它们各自表达行使新的功能。（B）由于功能的重叠，一个拷贝的突变（红色十字）只能揭示其一部分功能；只有两个拷贝的双突变才能说明此基因行使的全部功能。同样的原理也适用于在同一位置（如在单细胞物种中）的重复基因，它们一同作用或者分别在不同的环境中发挥作用。因此，基因重复复杂了所有物种中的遗传分析

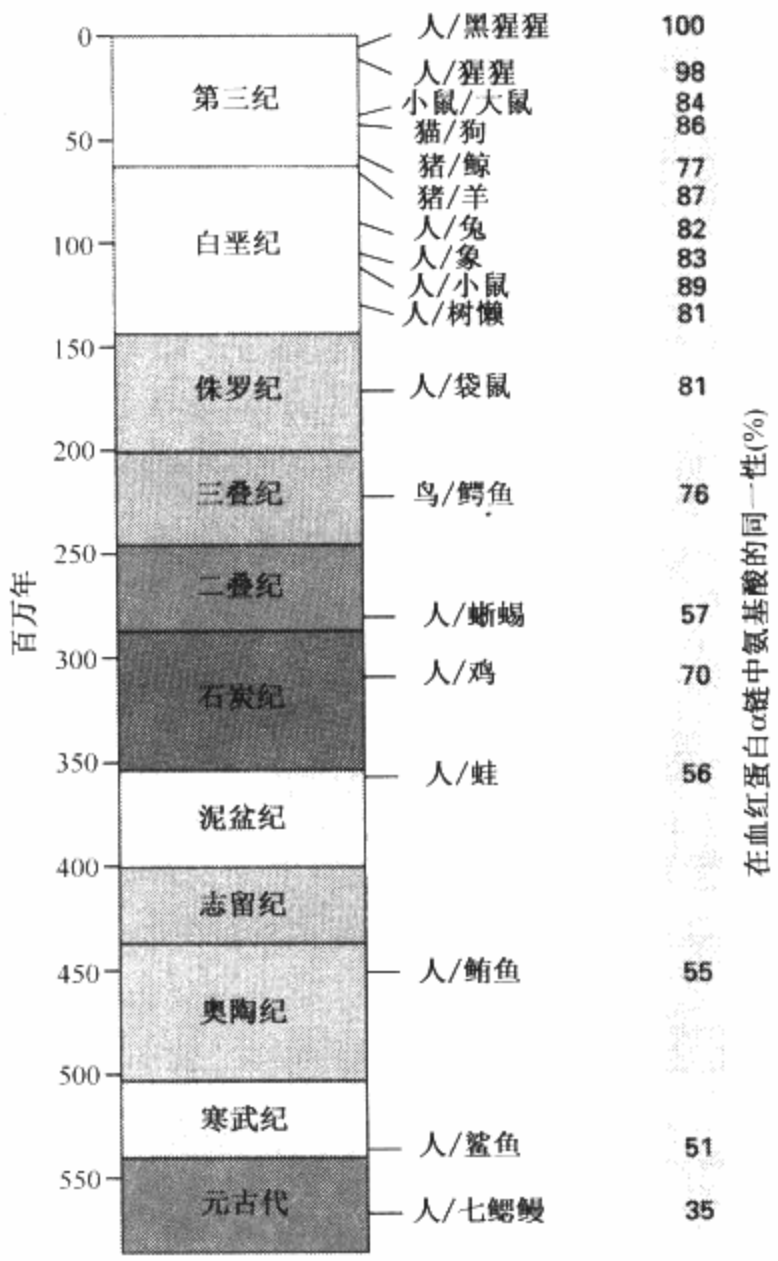


图 1-52　不同脊椎动物的分化时间

左边的刻度代表每对动物上一个共同祖先存在的估计时间以及地质年代。估计时间是根据正向同源蛋白质氨基酸序列的比对建立的；一对动物独自进化的时间越长，氨基酸的同源性越低。对于多种不同蛋白质数据取平均值得到最终的估计时间，时间刻度也已按照化石证据进行标记。例如，哺乳动物和鸟类的最后一个共同祖先出现在 3.1 亿年前。右边的图表示了一个特定蛋白质（任意选择）——血红蛋白质的 α 链的序列差异度。注意虽然这种蛋白质的总体趋势是随着时间的增大其差异度也增大，但也有一些不规则的地方。这些反映了进化过程的随机性，也有可能是自然选择使得一些有着特殊生理需求的物种的血红蛋白质序列改变速度加快。平均来说，在任何一个特定进化谱系中，血红蛋白质每一亿年在 100 个氨基酸中会有 6 个发生改变。一些有着严格功能限制的蛋白质，进化速度要慢得多，还有一些蛋白质进化速度是血红蛋白质的 5 倍。所有的这些都给分化时间的估计带来了不确定性，一些专家认为哺乳动物的分化时间要比所示的晚 6000 万年（摘自 S. Kumar and S. B. Hedges, *Nature* 392: 917～920, 1998）

基因组重复为复杂生物形式的出现提供了可能；它使物种有着丰富的闲置基因，可以为特殊的演化目的进行突变。有可能一个拷贝富集于肝脏中，而另一个富集于脑中发挥另一项功能。冗余基因按照这种方式制造复杂性和多变性。当这些基因执行多种功能时，就不会显得冗余了。然而，通常当基因执行特殊功能时，它们还必须同时执行自身的核心功能。单个基因的突变只能引起相对较小的变化，仅揭示基因的一小部分功能（图 1-51）。在脊椎动物的分子生物学中，普遍存在有着多种功能的基因家族功能的重叠现象，这些我们以后还会再提到。

小鼠是哺乳动物的模式生物

哺乳动物的基因数量一般是果蝇的 3 或 4 倍，基因组是其 20 倍，成体细胞数量是其百万到亿万倍。就基因组大小、功能、细胞生物学以及分子机制来说，哺乳动物是生命体的高级形式。从解剖学上讲，哺乳动物间主要是大小的差别，很难发现人体缺少某种大象或小鼠中的器官，反之亦然。进化任意地改变了外形，但是它无法改变结构上的逻辑性。

如果要得到两种哺乳动物种类之间确切的遗传差异，我们可以对其相应（或同源的）基因或其编码的氨基酸序列进行比对。其结果千差万别。但一般认为，如果我们对某种人类蛋白质和其他物种的同源蛋白质比对进行排序的话，那么，一种人类蛋白质和大象中同源蛋白质有 85％的相似性，和鸟类有 70％的相似性，差异性是前者的 2 倍，原因是鸟类和哺乳动物早于大象和人分支前 2 倍的时间就已产生了分化（图 1-52）。

小鼠由于体积小、易喂养，所以成了研究脊椎动物分子遗传学最初的模式物种。人们现已得到了许多自然突变的小鼠，弥补了人类缺乏相应的突变（图 1-53）。而且现在也产生了通过制造突变体来测定某种小鼠基因或基因组非编码区功能的方法，这些我们下面将会讲到。

图 1-53　人类和小鼠：相似的基因，相似的发育过程

图示的小孩和小鼠额头有着相似的白斑，都是由于同一个基因（叫做 *kit*）的突变造成的，此基因是色素细胞发育过程中所必需的（摘自 R. A. Fleischman，*Proc. Natl. Acad. Sci. USA* 88：10885～10889，1991. © National Academy of Sciences）

特殊定做的突变体小鼠为细胞生物学家们提供了宝贵的信息。它们为我们揭示了特定突变对于不同背景宿主的影响，同时验证了基因在被影响到的不同种类细胞内的作用。

人类可以报道自身的特性

当然，人们对于自身的基因组有着特殊的兴趣。我们希望了解构成自身的所有部分，并了解它们是怎么工作的。但是，即使你是一只小鼠，已经弄清楚了自身的分子生物学机制，人类依旧是一个诱人的模式遗传学物种，因为他具有以下特质：通过医学检查和自我报告，人们能够把自身的遗传（或其他的）紊乱编目。人类的种群是庞大的，目前已有 60 亿人口，这种自存档的特点意味着人类突变体有一个庞大的信息库。人类全基因组最近刚得到草图，有超过 30 亿个碱基，这使得人们可以更容易地精细研究何种基因导致人类的突变特征。

我们大体了解人类、小鼠、果蝇、线虫、酵母、植物以及细菌，通过基因序列比对勾画出两种模式物种的相关性，我们可以拓展对于它们所有物种的了解程度。

精确来说我们所有人都是不同的

人类基因组精确来说是什么呢？谁的基因组？平均来算，两个不同的人类个体 DNA 序列上每 1000 个碱基就会有一到两个不同。人类基因组计划从一小部分匿名者中随意选取 DNA 进行测序。人类基因组（整个人类的基因组）确切地来说，是一个更为复杂的概念，包括各种各样的已经存在于人类之中或者继续通过有性生殖交换和重排的基因集合。我们期望最终也可以记录这些变化。例如，对于这些知识的掌握可以让我们了解为什么一部分人易于感染某种疾病，另一部分对另一种疾病免疫力较低；为什么一些人对于某种药物成良性反应，而另一部分呈不良反应。这些还能为我们的历史研究提供更新的线索，研究我们祖先的种群迁移行为，他们所患的传染病，以及他们的日常饮食。所有这些都在人类社会的基因中留下了痕迹。

这些知识为我们干涉自然界的行为提供了动力，对于人类，可以避免或者预防疾病；对于植物，可以产生更好的植株；对于细菌，可以为我们所用。由于所有物种的遗传信息都以一种语言存储的，所以上述这些生物学事件都是相关的。分子生物学家们通读以及译码遗传信息的能力已经开始改变我们这个世界的各种关系。下一章节我们将介绍细胞生物学，我们

希望你们能够理解这项科学，有可能的话请投入到21世纪这项伟大的科学事业中来。

小结

真核细胞可以把它们的DNA包裹到一个独立的膜包裹单元——细胞核中。它们有负责移动的细胞骨架，可以进行消化和分泌的精细细胞内结构，吞噬其他细胞的功能（许多物种中），以及依赖于线粒体的氧化机制。

参考文献

一般文献

Alberts B, Bray D, Johnson A et al. (1997) Essential Cell Biology. London: Garland Publishing.

Darwin C (1859) On the Origin of Species. London: Murray.

Graur D & Li W-H (1999) Fundamentals of Molecular Evolution, 2nd edn. Sunderland, MA: Sinauer Associates.

Madigan MT, Martinko JM & Parker J (2000) Brock's Biology of Microorganisms, 9th edn. Englewood Cliffs, NJ: Prentice Hall.

Margulis L & Schwartz KV (1998) Five Kingdoms: An Illustrated Guide to the Phyla of Life on Earth, 3rd edn. New York: Freeman.

Watson JD, Hopkins NH, Roberts JW et al. (1987) Molecular Biology of the Gene, 4th edn. Menlo Park, CA: Benjamin-Cummings.

The Universal Features of Cells on Earth

Brenner S, Jacob F & Meselson M (1961) An unstable intermediate carrying information from genes to ribosomes for protein synthesis. *Nature* 190, 576–581.

Fraser CM, Gocayne JD, White O et al. (1995) The minimal gene complement of *Mycoplasma genitalium*. *Science* 270, 397–403.

Koonin EV (2000) How many genes can make a cell: the minimal-gene-set concept. *Annu. Rev. Genom. Hum. Genet.* 1, 99–116.

Watson JD & Crick FHC (1953) Molecular structure of nucleic acids. A structure for deoxyribose nucleic acid. *Nature* 171, 737–738.

Yusupov MM, Yusupova GZ, Baucom A et al. (2001) Crystal structure of the ribosome at 5.5Å resolution. *Science* 292, 883–896.

The Diversity of Genomes and the Tree of Life

Blattner FR, Plunkett G, Bloch CA et al. (1997) The complete genome sequence of *Escherichia coli* K-12. *Science* 277, 1453–1474.

Cole ST, Brosch R, Parkhill J et al. (1998) Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence. *Nature* 393, 537–544.

Dixon B (1994) Power Unseen: How Microbes Rule the World. Oxford: Freeman.

Doolittle RF (1998) Microbial genomes opened up. *Nature* 392, 339–342.

Doolittle WF (1999) Phylogenetic classification and the universal tree. *Science* 284, 2124–2129.

Henikoff S, Greene EA, Pietrokovski S et al. (1997) Gene families: the taxonomy of protein paralogs and chimeras. *Science* 278, 609–614.

Kerr RA (1997) Life goes to extremes in the deep earth—and elsewhere? *Science* 276, 703–704.

Koonin EV (2001) Computational genomics. *Curr. Biol.* 11, R155–R158.

Koonin EV, Aravind L & Kondrashov AS (2000) The impact of comparative genomics on our understanding of evolution. *Cell* 101, 573–576.

Mayr E (1998) Two empires or three? *Proc. Natl. Acad. Sci. USA* 95, 9720–9723.

Ochman H, Lawrence JG & Groisman EA (2000) Lateral gene transfer and the nature of bacterial innovation. *Nature* 405, 299–304.

Olsen GJ & Woese CR (1997) Archaeal genomics: an overview. *Cell* 89, 991–994.

Pace NR (1997) A molecular view of microbial diversity and the biosphere. *Science* 276, 734–740.

Tatusov RL, Koonin EV & Lipman DJ (1997) A genomic perspective on protein families. *Science* 278, 631–637.

The Institute of Genome Research (TIGR) Gene Indices. http://www.tigr.org.tdb/tgi.shtml

Woese C (1998) Default taxonomy: Ernst Mayr's view of the microbial world. *Proc. Natl. Acad. Sci. USA* 95, 11043–11046.

Woese C (1998) The universal ancestor. *Proc. Natl. Acad. Sci. USA* 95, 6854–6859.

Genetic Information in Eucaryotes

Adams MD, Celniker SE, Holt RA et al. (2000) The genome sequence of *Drosophila melanogaster*. *Science* 287, 2185–2195.

Andersson SG, Zomorodipour A, Andersson JO et al. (1998) The genome sequence of Rickettsia prowazekii and the origin of mitochondria. *Nature* 396, 133–140.

Arabidopsis Genome Sequencing Consortium (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* 408, 796–815.

Carroll SB, Grenier JK & Weatherbee SD (2001) From DNA to Diversity: Molecular Genetics and the Evolution of Animal Design. Maldon, MA: Blackwell Science.

Cavalier-Smith T (1987) The origin of eukaryote and archaebacterial cells. *Ann. NY Acad. Sci.* 503, 17–54.

DeRisi JL, Iyer VR & Brown PO (1997) Exploring the metabolic and genetic control of gene expression on a genomic scale. *Science* 278, 680–686.

Elgar G, Sandford R, Aparicio S et al. (1996) Small is beautiful: comparative genomics with the pufferfish (Fugu rubripes). *Trends Genet.* 12, 145–150.

Goffeau A, Barrell BG, Bussey H et al. (1996) Life with 6000 genes. *Science* 274, 546–567.

Gray MW, Burger G & Lang BF (1999) Mitochondrial evolution. *Science* 283, 1476–1481.

Holland PW (1999) Gene duplication: past, present and future. *Semin. Cell Dev. Biol.* 10, 541–547.

International Human Genome Sequencing Consortium (2001) Initial sequencing and analysis of the human genome. *Nature* 409, 860–921.

Lynch M & Conery JS (2000) The evolutionary fate and consequences of duplicate genes. *Science* 290, 1151–1155.

NCBI Genome Guide. http://www.ncbi.nlm.nih.gov/genome/guide/human/

Online Mendelian Inheritance in Man. http://www.ncbi.nlm.nih.gov/omim

Owens K & King M-C (1999) Genomic views of human history. *Science* 286, 451–453.

Palmer JD & Delwiche CF (1996) Second-hand chloroplasts and the case of the disappearing nucleus. *Proc. Natl. Acad. Sci. USA* 93, 7432–7435.

Philippe H, Germot A & Moreira D (2000) The new phylogeny of eukaryotes. *Curr. Opin. Genet. Dev.* 10, 596–601.

Plasterk RH (1999) The year of the worm. *Bioessays* 21, 105–109.

Rubin GM, Yandell MD, Wortman JR et al. (2000) Comparative genomics of the eukaryotes. *Science* 287, 2204–2215.

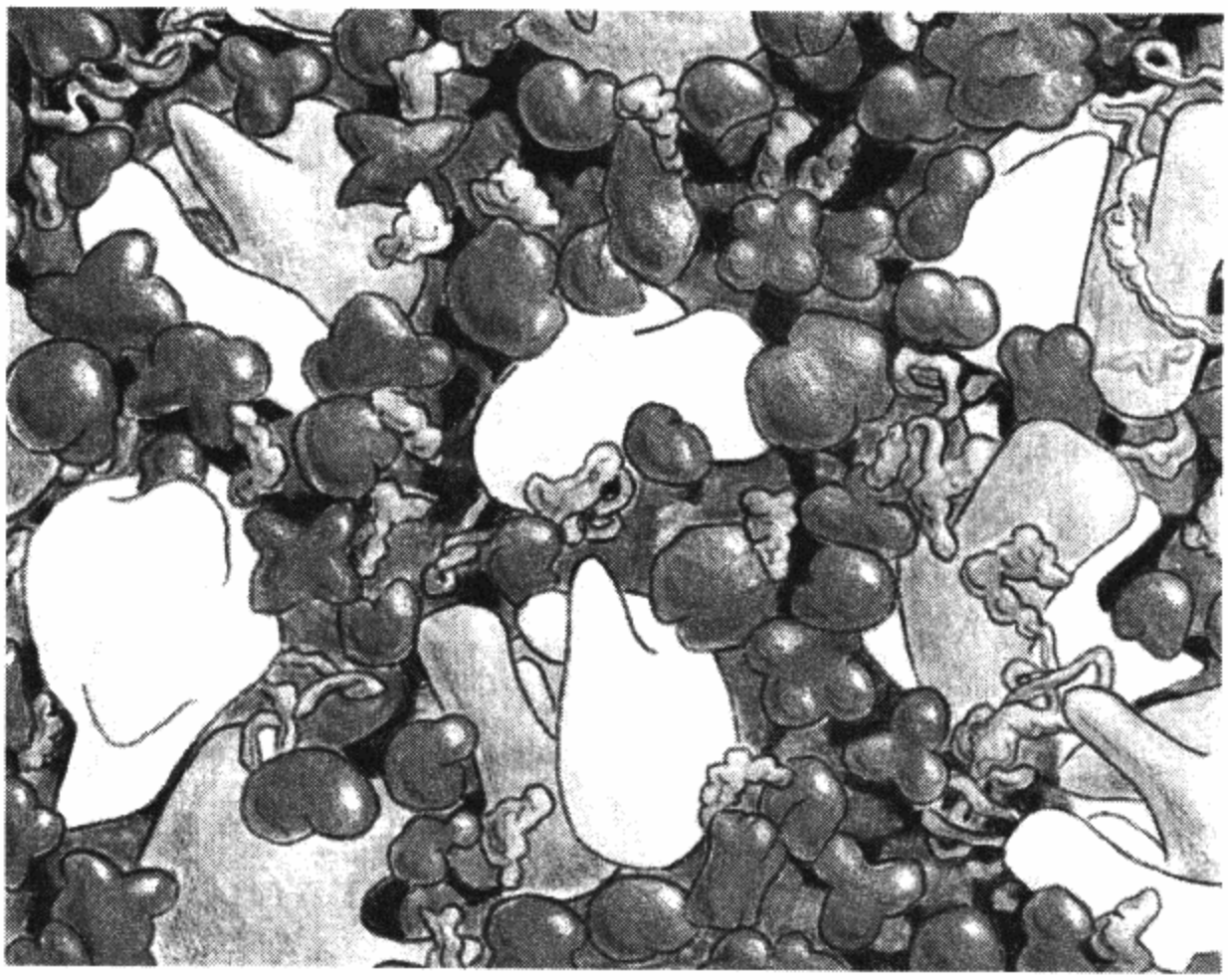
Sturtevant AH & Beadle GW (1939) An Introduction to Genetics. Philadelphia: Saunders. Reprinted 1988 New York: Garland.

The C. elegans Sequencing Consortium (1998) Genome sequence of the nematode *C. elegans*: a platform for investigating biology. *Science* 282, 2012–2018.

Tinsley RC & Kobel HR (eds) (1996) The Biology of *Xenopus*, p 16, pp 379, pp 391–401. Oxford: Clarendon Press.

Venter JC, Adams MD, Myers EW et al. (2001) The sequence of the human genome. *Science* 291, 1304–1351.

Walbot V (2000) *Arabidopsis thaliana* genome. A green chapter in the book of life. *Nature* 408, 794–795.



拥挤的细胞质。这幅只显示大分子物质的按照原比例放大的图像，给予我们一个较好的有关细胞质是非常拥挤的印象。蓝色的是 RNA，绿色的是核糖体，红色的是蛋白质（改编自 D. S. Goodsell，*Trends Biochem. Sci.* 16：203～206. 1991.）

2　细胞的化学和生物合成

细胞的化学组成
催化作用和细胞利用能量
细胞怎样从食物中获取能量

乍一看，我们很难相信上一章所介绍的每种生物都只不过是一个生物系统。很多种生命形式，它们有功能的行为以及它们生长和繁殖的能力都似乎不同于化学上通常所介绍的固体、液体和气体世界。事实上，直到 19 世纪人们还认为动物含有活力元素——泛灵素，独特地负责动物的不同性质。

我们现在知道不存在违背化学和物理规律的生物体。然而，生命化学事实上是一种特殊的化学。第一，生命化学主要以碳化合物为基础，而碳化合物的研究称为有机化学。第二，细胞含有 70%的水分。生命几乎完全依赖于水溶液里的化学反应。第三，最重要的是，细胞化学非常复杂。即使最简单的细胞在化学方面比任何其他已知的化学系统要复杂得多。尽管细胞含有许多种小的含碳分子，细胞中大部分碳原子被合并成巨大的多聚体分子——头尾相接的化学亚基链。正是这些大分子的独特性质才使得细胞和生物体生长和繁殖，同时做生命所特有的所有其他事情。

细胞的化学组成

物质是由元素组合而成，元素是指不能通过化学手段分解或转化成其他物质的物质，如氯或碳。元素中能保持其特有化学性质的最小颗粒是原子。然而，除了单纯元素外，物质（包括形成活细胞的材料）的特征取决于其原子一起成群连接以形成分子的方式。因此，为了理解生物是如何由无机物质构建的，关键是要知道所有的化学键是如何形成的，化学键使分子中的原子聚集在一起。

细胞是由几种原子构成的

每个**原子**（atom）的中央有一个带正电荷的核，在离核的一定距离处，核被一层带负电荷的**电子**（electron）环绕，这些电子通过核的静电吸引被固定在一系列的轨道上。核是由两种亚原子颗粒构成的：带正电荷的**质子**（proton）和不带电荷的中子。原子核的质子数决定原子序数。氢原子的原子核只有一个质子，所以原子序数为 1 的氢是最轻的元素。碳原子的核有 6 个质子，它的原子序数则为 6（图 2-1）。每个质子所带电荷正好与每个电子的电荷数相等，电性相反。因为每个原子是电中性的，所以围绕核的电子数与核质子数相等。因此，一个原子的电子数也等于原子序数。正是这些电子决定了原子的化学性质，一个指定元素的所有原子都具有相同的原子序数。

中子（neutron）是质量与质子基本相同的不带电荷的亚原子颗粒。中子有助于稳定核的结构——如果中子数过多或过少的话，核就会经放射性衰变而裂变。但中子并不改变原子的

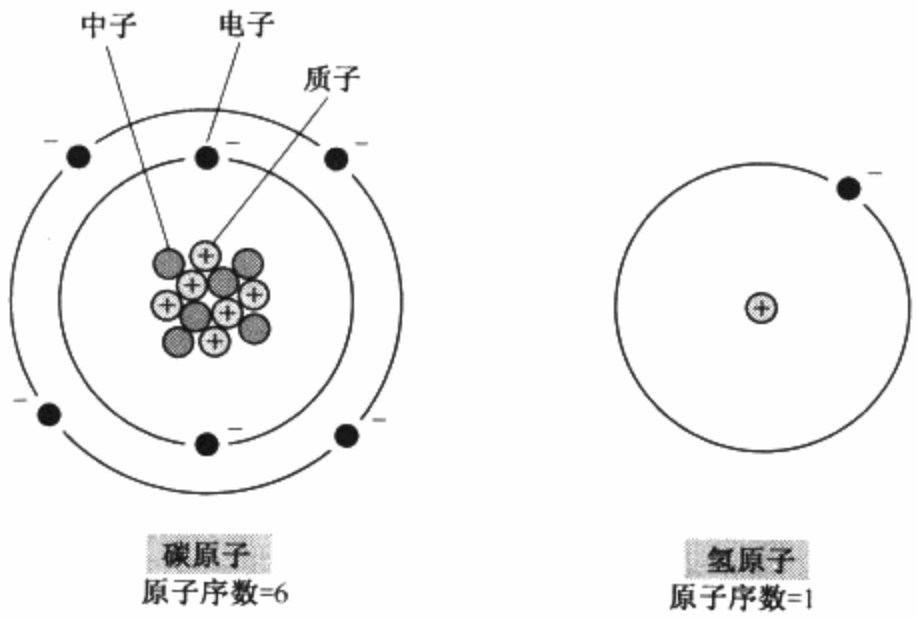


图 2-1　一个碳原子和一个氢原子的高度图解表示法

尽管图中的电子以单个颗粒标出，事实上它们的行为由量子力学定律控制的，无法准确地预测在任何指定时刻电子的位置。除了氢原子外，每一个原子的核是由带正电荷的质子和电中性的中子组成。一个原子的电子数等于其质子数（原子数），所以原子不带电荷。事实上，中子、质子和电子相对于原子整体而言是微小的，它们的大小在图中被大大地夸大了。此外，核的直径只有电子云的大约 10^{-4}

化学性质。因此，元素能以几种物理性质不同而化学形式相同的形式存在，这些形式的元素称为同位素。每种同位素的中子数不同，而原子数相同。自然界几乎存在所有元素的多种同位素，包括一些不稳定的元素。例如，地球上存在的大多数碳是稳定的同位素碳 12（6 个质子和 6 个中子），但也有少量的不稳定同位素，即放射性的碳 14（6 个质子和 8 个中子）。碳 14 以缓慢而稳定的速率发生放射性衰变。这是被称为碳 14 年代鉴定技术的基础。该技术在考古学中被用来确定有机物质的起源时间。

一个原子的**原子质量**（atomic weight），或分子的**分子质量**（molecular weight）是该原子相对于氢原子的质量。因为电子非常轻，可以忽略不计，所以原子质量或分子质量基本等于该原子或分子所含有的原子和中子数。因此碳的主要同位素的原子质量是 12，标记为^{12}C，而刚谈过的不稳定同位素的原子质量是 14，写为^{14}C。原子或分子的质量常以道尔顿（Da）为单位，1Da 即一个原子的质量单位，大约等于一个氢原子的质量。

原子非常小，以至于无法想像其大小。单个碳原子的直径大于 0.2nm，将 500 万个碳原子排成一条直线才 1mm。一个质子或中子大约重$1/(6\times10^{23})$g，所以 1g 氢含有 6×10^{23} 个原子。这个巨大数字 6×10^{23}，称为**阿伏加德罗常数**（Avogadro's number），它是描述常规质量和单个原子或分子质量的关键标度因子。如果一种物质的分子质量为 X，则 6×10^{23} 个分子的该物质的质量为 Xg。该数量称为 1mol 的该物质（图 2-2）。

一摩尔是Xg物质，X是其相对分子量(分子质量)。1mol含有6×10^{23}个分子的该物质。

1mol碳重12g,
一摩尔葡萄糖重180g
一摩尔氯化钠重58g。

摩尔溶液具有1L溶液含有1mol物质的浓度。例如，一摩尔溶液(表示为1mol)的葡萄糖具有180mg/L，而一毫摩尔溶液(1mmol)具有180mg/L。

克的标准缩写是g；升的缩写是L

图 2-2　摩尔和摩尔溶液

自然界中存在 92 种元素，每种元素的原子都含有不同于其他元素的质子数和电子数。然而，生物体只由少数元素构成，其中 4 种，碳（C）、氢（H）、氮（N）和氧（O）占生物体

重量的 96.5%。这种组成完全不同于非生物的无机环境（图 2-3），同时也是一种特殊的化学证据。表 2-1 中列有生物体内最常见元素和一些原子特征。

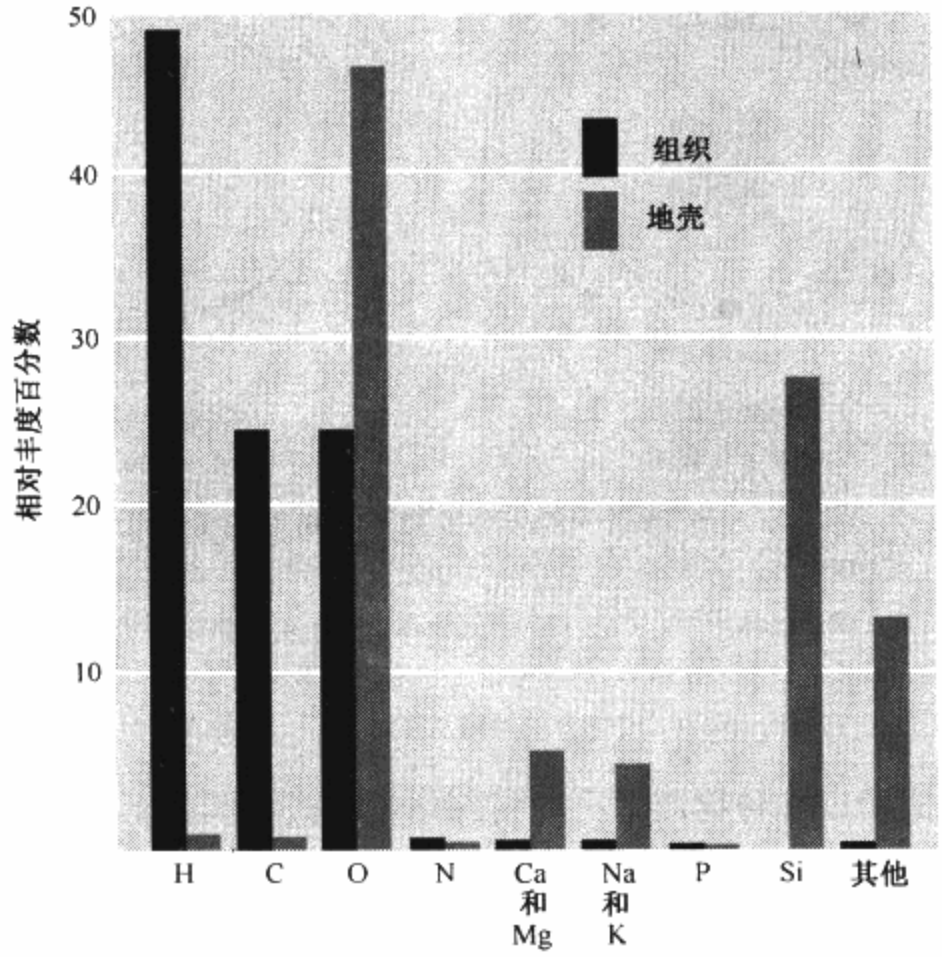


图 2-3　某些化学元素在非生命世界（地壳）中的丰度与它们在动物组织中的丰度的比较

每一种元素的丰度以占样品中存在的原子总数的百分比表示。因此，生物体中约 50%的原子是氢原子。图中的概括排除了矿化组织（如骨和牙齿），因为它们含有大量的无机钙盐和磷。元素的相对丰度在所有生物体中是相似的

表 2-1　活组织内含量最丰富的元素的原子特征

生物组织中的常见元素						
元素		质子	中子	电子	原子序数	原子质量
氢	H	1	0	1	1	1
碳	C	6	6	6	6	12
氮	N	7	7	7	7	14
氧	O	8	8	8	8	16
稀有元素						
元素		质子	中子	电子	原子序数	原子质量
钠	Na	11	12	11	11	23
镁	Mg	12	12	12	12	24
磷	P	15	16	15	15	31
硫	S	16	16	16	16	32
氯	Cl	17	18	17	17	35
钾	K	19	20	19	19	39
钙	Ca	20	20	20	20	40

最外层电子决定原子的作用方式

为了理解原子如何结合在一起形成分子（分子形成生物体），我们必须特别注意电子。核中质子和中子互相紧密连接，只有在极端条件（如放射性衰变或太阳内部或核反应堆内部）才改变其配体。在活组织中，只有原子的电子才进行重排。电子形成原子的外部，它们决定与哪些原子结合形成分子。

电子围绕着核不停地运动，但这种亚显微水平的运动遵循不同于我们日常生活所熟悉的法规。这些法则规定原子中的电子只存在于被称为轨道的某些不连续的状态，某个特定的轨道只能容纳有限的电子，即所谓的电子层。

一般最靠近带正电荷的核的电子与核之间的吸引力最强，占据最里且结合最紧的电子层，该电子层最多能容纳 2 个电子、第二电子层离核远些，电子的结合也更松些，最多能容纳 8 个电子。第三电子层的电子结合更松些，也最多能容纳 8 个电子。第四和第五电子层均能容纳 18 个电子。生物分子中很少含有多于 4 个电子层的原子。

当所有电子均处于结合最紧密的状态（即占据最内层）时，原子的电子排列最稳定。因此，除了更大原子的某些例外，原子的电子按顺序充满轨道——第一电子层、第二电子层、第三电子层等。最外电子层完全充满电子的原子特别稳定，因此不发生化学反应。这些原子有氦（2 个电子）、氖（2＋8 个电子）和氩（2＋8＋8 个电子），这些都是惰性气体。然而，氢只有 1 个电子，因此其电子层为半充满，其反应性很强。同样，活组织中其他原子都含有未充满的外电子层，因此能够相互给出、接受或共享电子，从而形成分子和离子（图 2-4）。

原子序数	元素	电子层 I	II	III	IV
1	氢	●			
2	氦	●●			
6	碳	●●	●●●●		
7	氮	●●	●●●●●		
8	氧	●●	●●●●●●		
10	氖	●●	●●●●●●●●		
11	钠	●●	●●●●●●●●	●	
12	镁	●●	●●●●●●●●	●●	
15	磷	●●	●●●●●●●●	●●●●●	
16	硫	●●	●●●●●●●●	●●●●●●	
17	氯	●●	●●●●●●●●	●●●●●●	
18	氩	●●	●●●●●●●●	●●●●●●●●	
19	钾	●●	●●●●●●●●	●●●●●●●●	●
20	钙	●●	●●●●●●●●	●●●●●●●●	●●

图 2-4　某些常见元素中充满和未充满的电子层

生物体内所有的常见元素具有未充满的最外电子层（红色），因此能参与和其他原子的化学反应。为了比较，图中标出了某些只具有充满的电子层（黄色）的元素；这些元素不具有化学反应性

因为未充满的电子层不如已充满的电子层稳定，含有未充满的外电子层很容易通过获得或丢失足够的电子和其他原子作用，从而获得充满的外电子层。这种电子交换可以通过原子之间的电子转移或原子之间的电子共享来实现。这两种策略产生了两种**化学键**（chemical bond）：离子键（一个原子向另一原子供给电子）和共价键（两个原子共用一对电子）（图 2-5）。电子对的共用经常是不平等的，存在着原子间的部分转移，这种中间策略产生了极性共价键。我们将在以后介绍极性共价键。

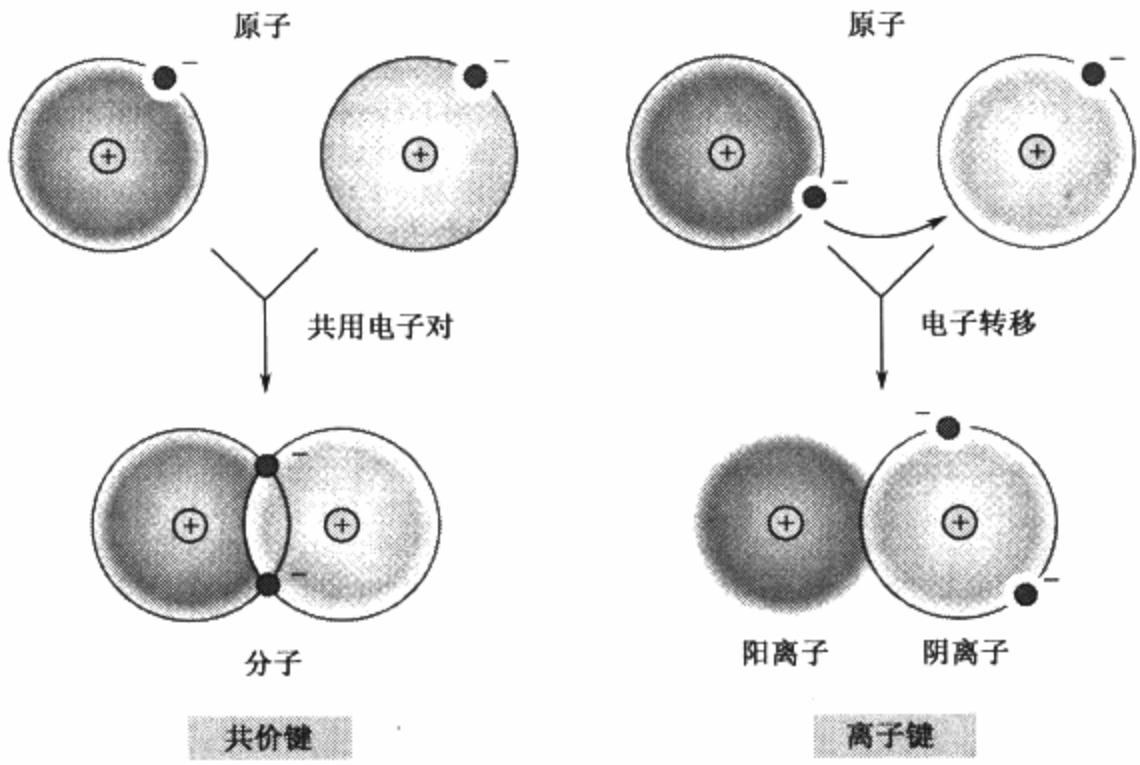


图 2-5　共价键和离子键的比较

原子能通过彼此的相互作用获得其最外电子层内更为稳定的电子排列。离子键是当电子从一个原子转移到另一个原子时形成的。共价键是当原子之间共用电子时形成的。图中所示的两个例子代表极端的离子；共价键经常是通过电子的部分转移（电子层的不均匀共用）形成，从而产生极性共价键（图 2-43）

氢原子只需一个电子就可以充满其电子层，一般通过电子共用获得另一个电子，从而与另一个原子形成一个共价键；在多数情况下，该共价键是极性的。活细胞中最常见的其他元素——C、N 和 O（具有未充满的第二电子层），以及 P 和 S（具有未充满的第三电子层）（图 2-4），通常通过形成几个共价键来共用电子和充满带 8 个电子的外电子层。为充满其外电子层，原子必须获得或丢失（共用或转移）的电子数被称为化合价。

外电子层在决定元素的化学性质中起着关键作用，这说明当原子按原子序数排列时，会周期性地出现具有相似性质的元素。例如，只含有一个电子的第二电子层的元素与第二电子层充满而带一个电子的第三电子层的元素性质相似。例如，金属的未充满电子层只含有一个或几个电子，而惰性气体含有充满的外电子层。

电子的获得和丢失形成离子键

除了已充满的外电子层还含有一个或两个电子的元素，或只需一个或两个电子就充满外电子层的元素最可能形成离子键。这些元素常常更容易通过与另一个原子之间的电子转移而不是电子共用获得完全充满的外电子层。例如，从图 2-4 我们可以看到钠（Na）原子（原子序数为 11）通过失去第二电子层外的一个电子能形成充满的电子层。而氯（Cl）原子（原子序数为 17）通过获得一个电子，能充满其外电子层。因此，如果 Na 原子碰到 Cl 原子。Na 原子上的一个电子会转移到 Cl 原子上，从而使得两个原子都具有充满的外电子层。柔软而反应性强的金属钠和绿色的毒气氯结合的产物是食盐（NaCl）。

当电子从 Na 转移到 Cl 时，两个原子都成为带电荷的离子（ion）。失去一个电子的钠原子的电子比核中的质子少一个；因此，它带有一个正电荷（Na^+）。获得一个电子的 Cl 原子的电子比质子多一个，带有一个负电荷（Cl^-）。带正电荷的离子称为阳离子，带负电荷的离子称为阴离子。按照失去或得到的电子数，离子可进一步划分。钠（Na）和钾（K）可以丢失 1 个电子，从而可形成带 1 个正电荷的阳离子（Na^+ 和 K^+），而镁和钙可以丢失 2 个电子，

从而形成带 2 个正电荷的阳离子（Mg^{2+} 和 Ca^{2+}）。

因为 Na^+ 和 Cl^- 所带的电荷相反，它们互相吸引，因而以离子键（ionic bond）结合。盐晶体含有极大量的 Na^+ 和 Cl^-（1mm 宽的晶体中含有大约 2×10^{19} 个 Na^+ 和 Cl^-），这些 Na^+ 和 Cl^- 被一起填充在精确的三维排列中，与之相反的电荷被准确地平衡（图 2-6）。只靠离子键结合的物质（如 NaCl）通常称为盐，而不是分子。离子键只是存在于原子之间的几种非共价键中的一种，我们将会遇到其他的例子。

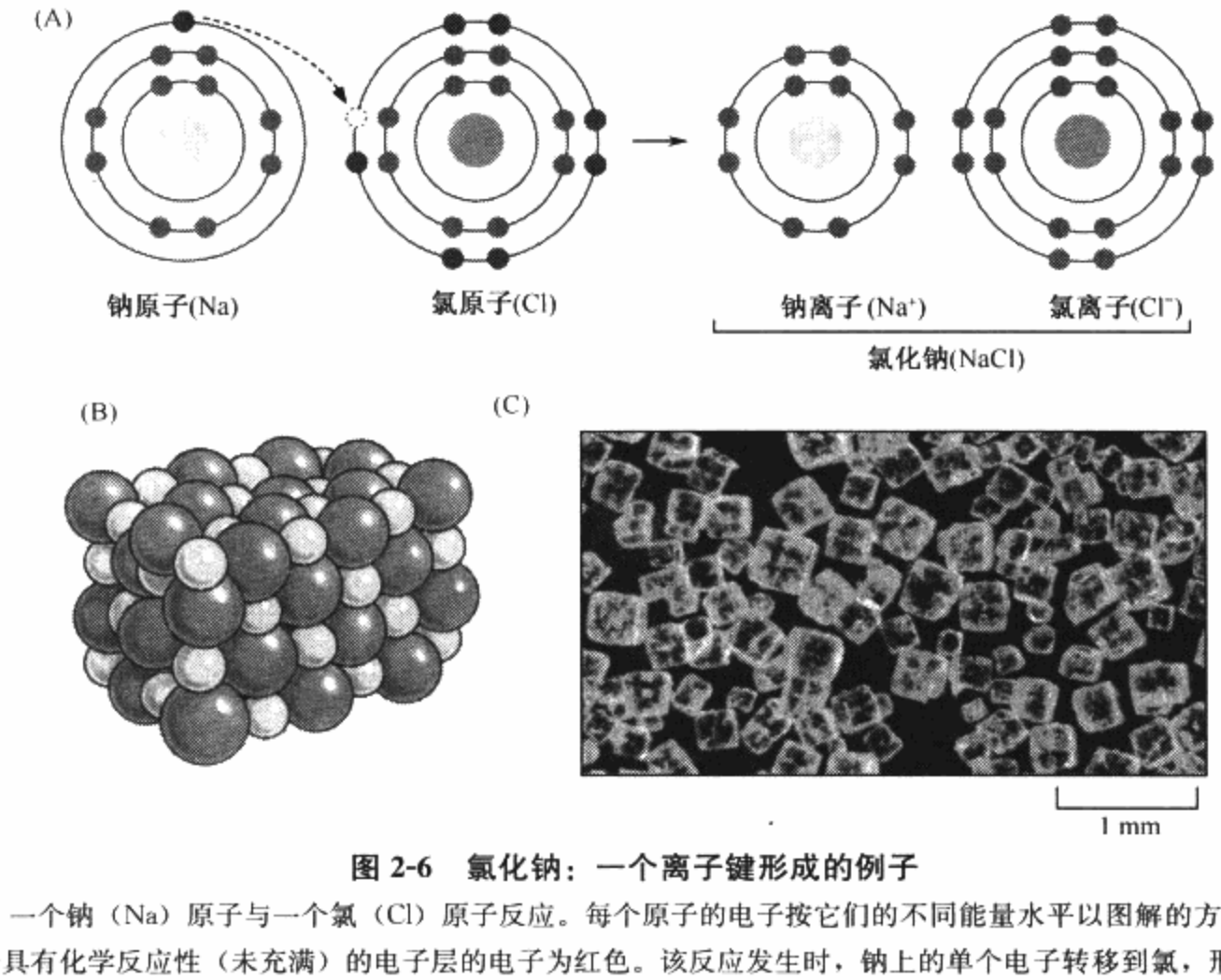


图 2-6　氯化钠：一个离子键形成的例子

（A）一个钠（Na）原子与一个氯（Cl）原子反应。每个原子的电子按它们的不同能量水平以图解的方式表示；位于具有化学反应性（未充满）的电子层的电子为红色。该反应发生时，钠上的单个电子转移到氯，形成两个带电荷的原子或离子，每个原子在其最外电子层具有充满的电子。带有相反电荷的两个离子通过静电吸引结合到一起。（B）钠和氯之间的反应产物，即结晶的氯化钠，是由钠离子和氯离子组成的，这些离子以规则的排列紧密填充在一起，在该排列中电荷被准确地平衡。（C）氯化钠晶体的照片

因为水和离子之间能发于有利的相互作用，水可大大地减弱离子键；因此许多盐（包括 NaCl）是高度水溶性的，解离成单个离子（如 Na^+ 和 Cl^-），每个离子被一群水分子围绕。相反，共价键的强度不受水的影响。

共价键是通过共用电子对形成的

细胞的所有特征取决于它所含有的分子。分子（molecule）是指由共价键（covalent bond）连接的一群原子。在共价键中，原子之间共用电子对，从而充满外电子层，而不是原子之间的电子转移。在可能的最简单分子——氢（H_2）分子中，2 个氢原子（每个带单个电子）共用 2 个电子，2 个电子是充满第一电子层所必需的数量。这些共用的电子形成带负电荷的电子云，该电子云在 2 个带正电荷的核之间最密，有助于 2 个核结合，与相似电荷之间的相互斥力相反，相互斥力迫使它们分开。当核被分开一个典型的距离时，吸引力和排斥力处于平衡，该距离称为键长。

任何键（共价或非共价）的另一个关键性质是它的强度。键能是指打断该键所必须提供的能量。键能经常用千卡每摩尔（kcal/mol）的单位来表示的，1kcal 是指将 1L 水升高 1℃所需要

的能量。因此，如果必须用 1kcal 的热量才能打断 6×10^{23} 个特异类型的键（即 1mol 的该键），那么该键的强度是 1kcal/mol。一个被广泛使用的等价能量单位是千焦耳（kJ），1kJ 等于 0.239kcal。

为了理解键能的含义，将键能与分子和其环境中其他分子不断进行碰撞所产生的平均能量（它们的热能），以及其他的生物能源（如光和葡萄糖的氧化）（图 2-7）作比较是有用的。典型的共价键比热能强 100 倍，所以共价键可以抵抗因热运动而产生的键断裂，共价键通常只在与其他的原子和分子发生的化学反应过程中被打断。共价键的生成和打断是剧烈的事件，在活细胞中它们受到高度特异的催化剂的严密控制，这些催化剂称为酶。非共价键通常非常弱，我们以后将看到在许多情况下非共价键在细胞中是重要的，在这些情况下分子必须容易结合和解离，从而执行它们的功能。

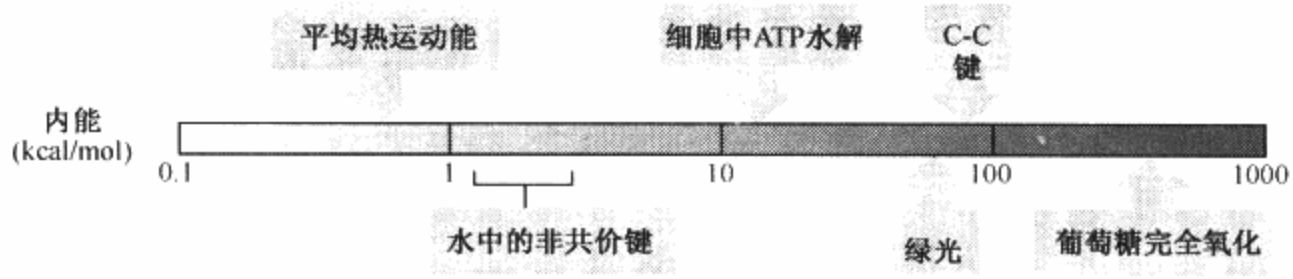


图 2-7　细胞的某些重要能量

注意这些能量是以对数标度比较的

一个 H 原子只能形成一个共价键，在细胞中形成共价键的其他常见原子，O、N、S 和 P 以及非常重要的 C 原子，能形成一个以上的共价键。正如我们已介绍的，这些原子的最外电子层最多能容纳 8 个电子，它们与许多其他必需原子形成共价键以达到 8 个电子。氧的外电子层带有 6 个电子，当通过与其他原子共用而获得另外 2 个电子时，是最稳定的，因此氧最多形成 2 个共价键。带有 5 个外电子的氮最多形成 3 个共价键，而带有 4 个外电子的碳最多形成 4 个共价键——共用 4 对电子（图 2-4）。

当一个原子与几个其他原子的形成共价键时，这些多键相互之间在空间上具有确切的方向，反映了共用电子的轨道方向。因此，多原子之间的共价键用特异的键角以及键长和键能描述（图 2-8）。例如，碳原子的周围能形成 4 个共价键，这些共价键的排列好像指向正四面体的 4 个角。共价键的精确定位是形成有机分子的三维几何的基础。

存在几种不同类型的共价键

大多数共价键需要共用 2 个电子，每个参与原子给出一个电子；这些共价键称为单键。然而，某些共价键需要共用一对以上的电子。例如，可以共用 4 个电子，每个参与原子给出 2 个电子；该键称为**双键**。双键比单键更短但更强，双键对含该双键的分子的三维几何具有典型的效应。2 个原子之间的单个共价键通常允许分子的一部分相对于该分子的另一部分绕键轴旋转。双键阻止这种旋转，产生更坚硬的和较不易变的原子排列（图 2-9 和专题 2-1）。

某些分子在 3 个或更多个原子之间共用电子，产生具有杂合特征的中间体，该中间体介于单键和双键之间。例如，高度稳定的苯分子是由 6 个碳原子的环组成的，在该碳环中键合电子平均分配于该碳环中（尽管苯分子通常被描述为单键和双键的间隔排列，如专题 2-1 中所示）。

当单个共价键连接的原子属于不同元素时，通常这两个原子吸引共用电子的程度不同。例如，与 C 原子相比，O 和 N 原子吸引电子相对强些，而 H 原子吸引电子更弱。**极性**（polar）结构（电子意义上）是指具有集中于一端的正电荷（正极）和集中于另一端的负电荷（负极）的结构。因此，电子以这种方式被不均匀地共用的共价键称为极性共价键（图 2-10）。

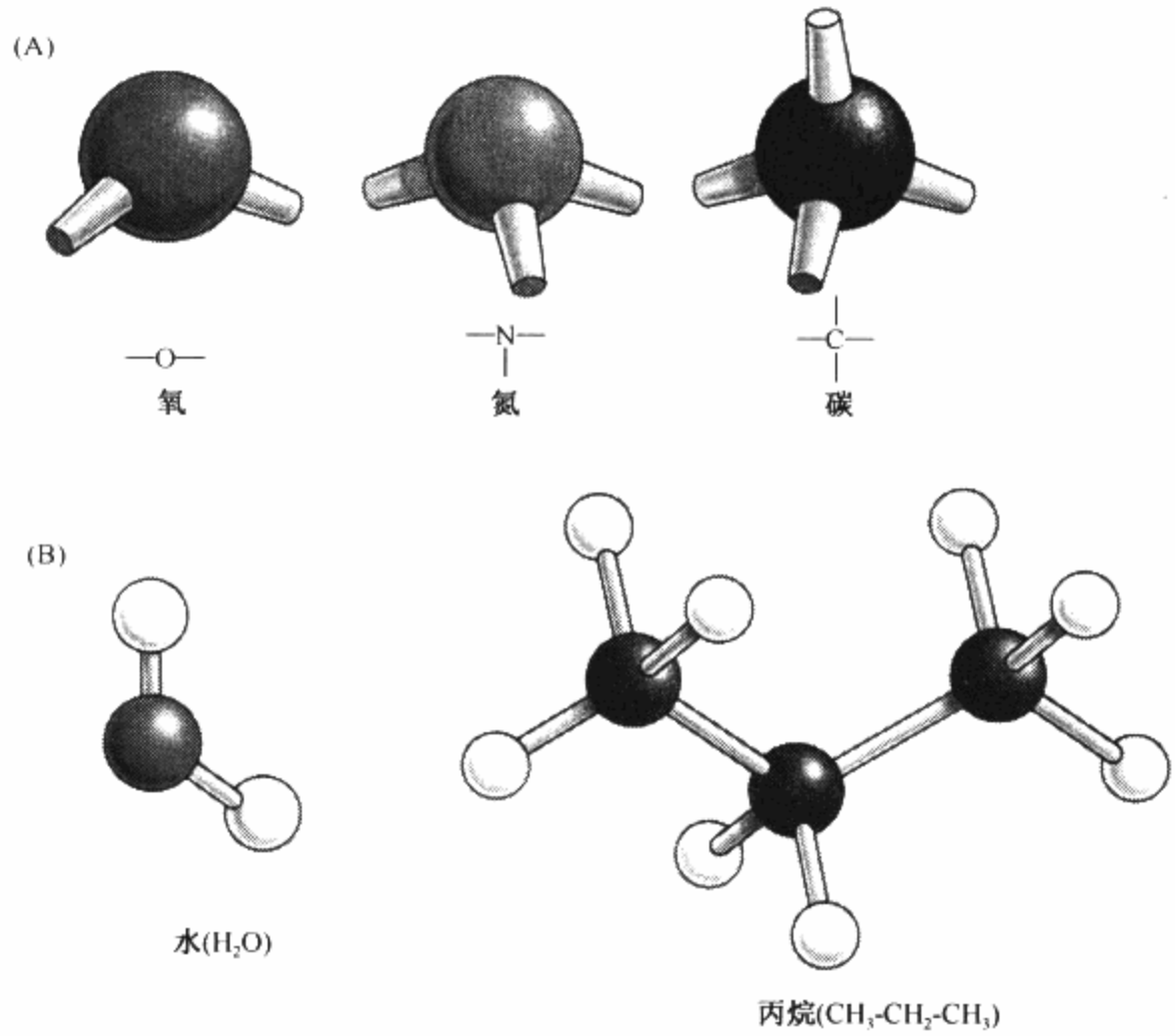


图 2-8　共价键的几何

（A）由氧、氮和碳形成的共价键的空间排列。（B）由这些原子形成的分子具有准确的三维结构，如图中水和丙烷的球棒模型所示。分子结构可以由每个共价连接的键角和键长来规定

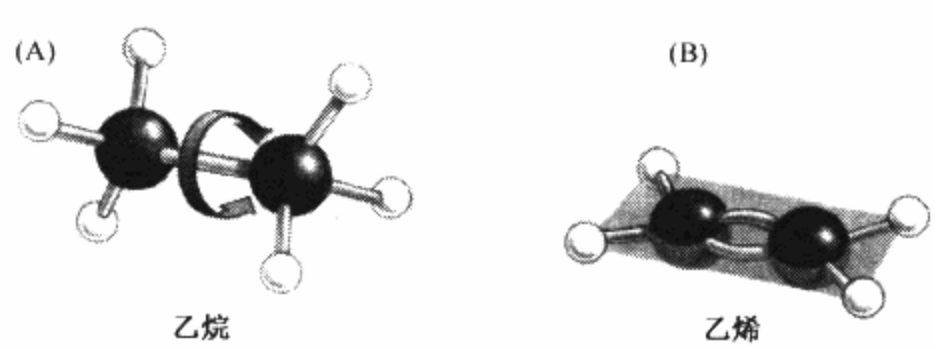


图 2-9　碳—碳双键和单键的比较

（A）乙烷分子在两个碳原子之间具有一个共价键，该图说明了共价单键的四面体排列。其中的一个由共价键连接的 CH_3 基团能沿键轴相对另一个 CH_3 基团旋转。（B）乙烯分子的两个碳原子之间的双键改变了碳原子的键几何，使所有的原子位于同一平面（蓝色）；双键阻止了一个 CH_2 基团相对于另一个 CH_2 基团的旋转

例如，氧和氢之间的共价键—O—H，或氮和氢之间的共价键—N—H，是极性的；而碳和氢之间的共价键—C—H 的电子被两个原子更均匀地吸引，因此是相对非极性的。

因为极性共价键形成永久偶极，而永久偶极使分子能通过电动力相互作用，所以极性共价键在生物学中极其重要。任何含有许多极性基团的大分子在其表面具有部分正电荷和负电荷的模式。当该分子遇到具有互补电荷的第二个分子时，这两个分子将通过永久偶极的相互作用而彼此吸引，这些永久偶极的相互作用像（但弱于）前面谈到的

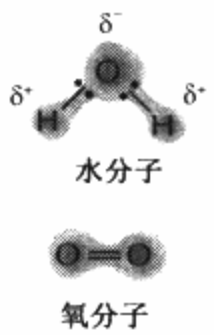


图 2-10　极性和非极性共价键

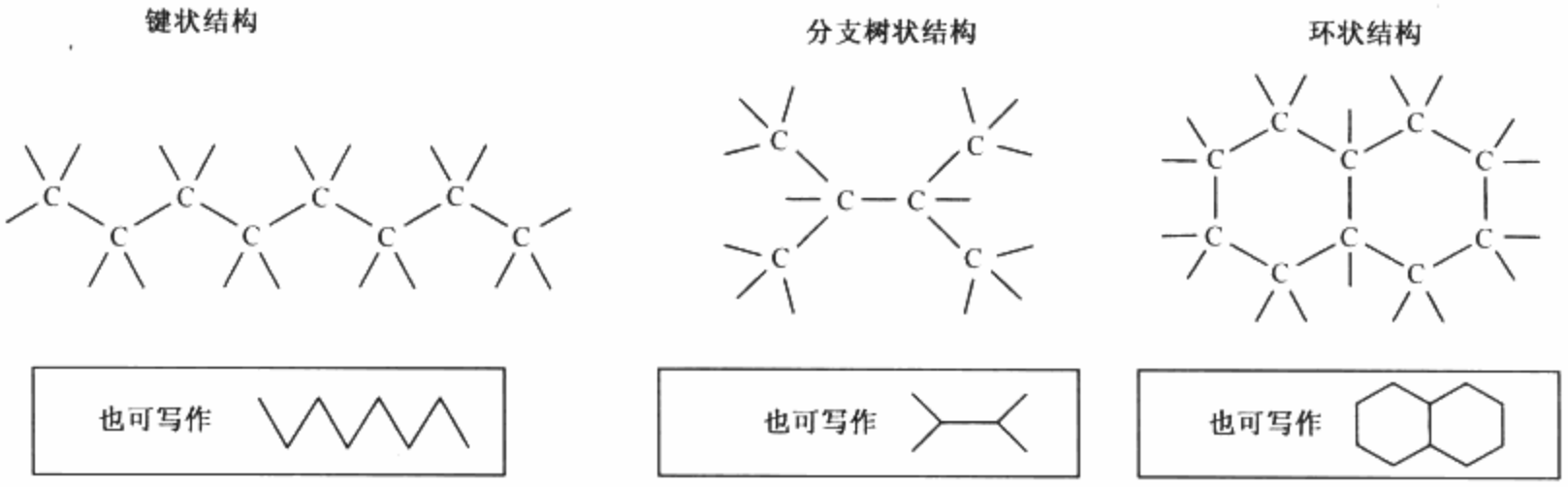
该图比较了极性水分子（H_2O）和非极性氧分子（O_2）中的电子分布（δ^+，部分正电荷，δ^-，部分负电荷）

NaCl 的离子键。

专题 2-1　生物分子中常出现的化学键和化学基团

碳骨架

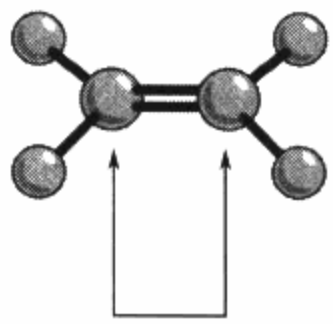
碳原子在细胞内发挥独特的作用，因为它能和其他的碳原子形成强硬的共价键。因此，碳原子可以连接起来形成链状结构，或分支树状结构，或环状结构。



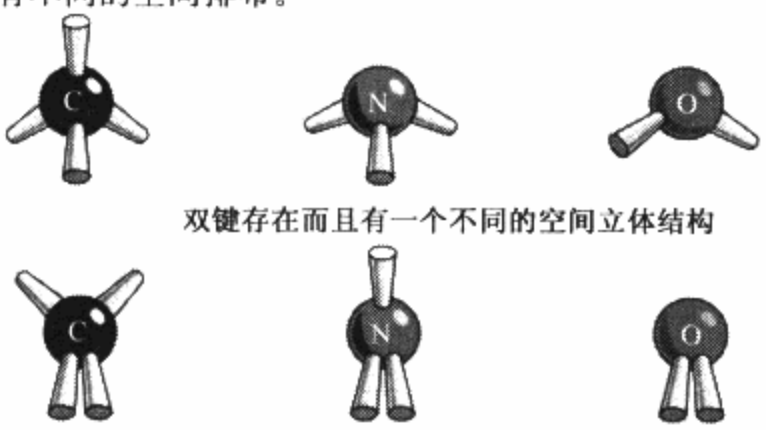
共价键

当 2 个原子彼此靠得很近，共用一个或多个电子时，便形成一个共价键。在单键中共用 2 个电子，双键中则共用 4 个电子。

每个原子在限定的空间排布中形成确定数目的共价键。例如，碳原子以四面体排布能形成 4 个单键，而氮原子形成 3 个单键，氧原子形成 2 个单键，如下所示双键存在并且具有不同的空间排布。

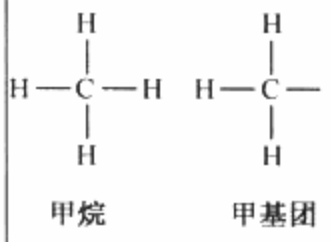


通过 2 个或更多共价键连接的原子不能绕键轴自由旋转。这种限制是影响许多大分子三维形状的主要因素之一。



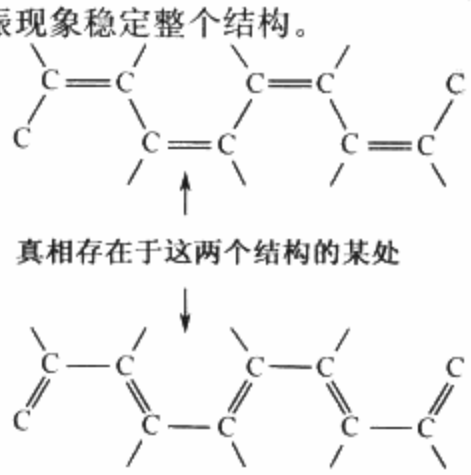
烃

碳原子和氢原子结合在一起形成稳定的复合物（或化学基团）叫做烃。烃没有极性，不形成氢键，通常不溶于水。

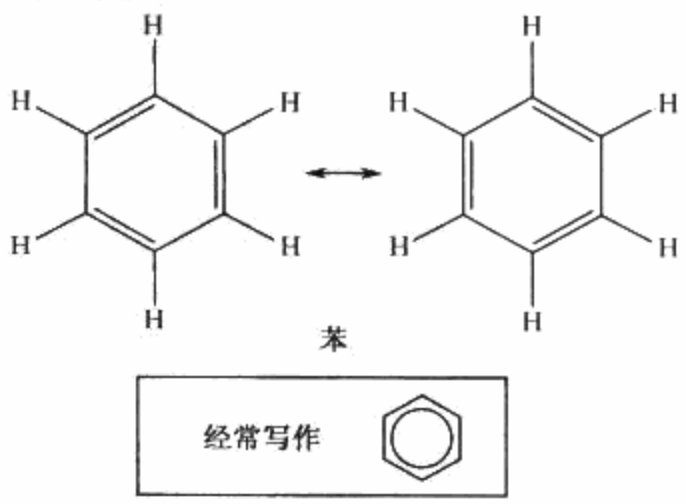


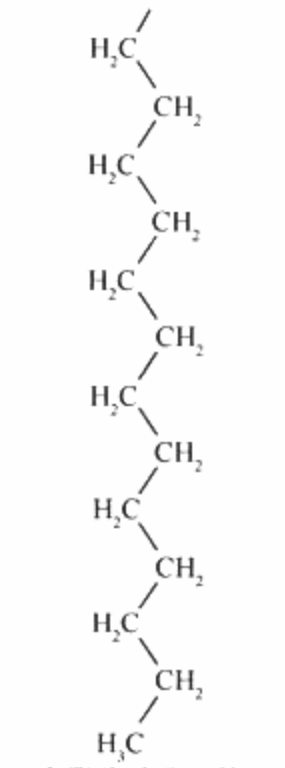
交替排列的双键

碳链可以含有双键。如果这些双键位于交替出现的碳原子上，双键中的电子就会在分子间移动，通过共振现象稳定整个结构。



环中交替排列的双键可以形成一个非常稳定的结构。

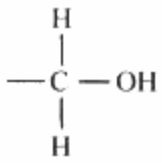




C—O 化学基团

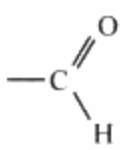
许多生物复合物含有碳氧键。例如，

醇

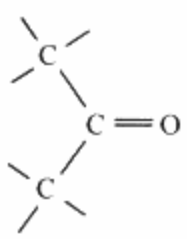


—OH 称为羟基。

醛

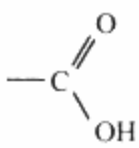


酮



C═O 称为羰基。

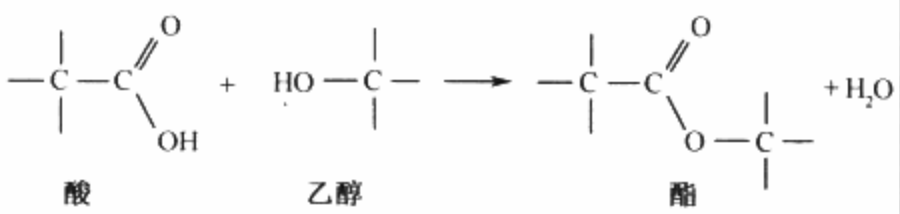
羧酸



—COOH 称为羧基。在水中，该基团失去一个 H^+ 称为—COO^-。

酯

酸与醇结合形成酯。

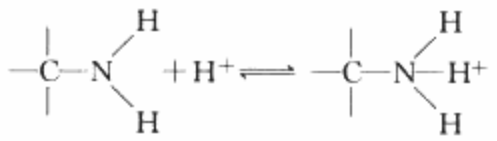


酸　　乙醇　　酯

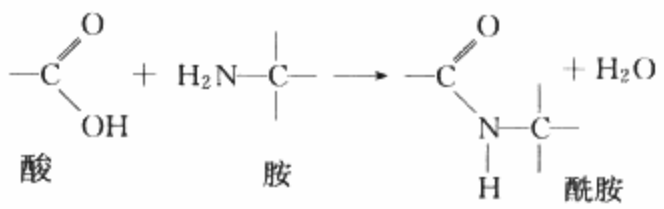
C—N 化学基团

胺和酰胺是两类重要的含碳氮键的复合物。

胺在水中结合一个 H^+，带正电荷。

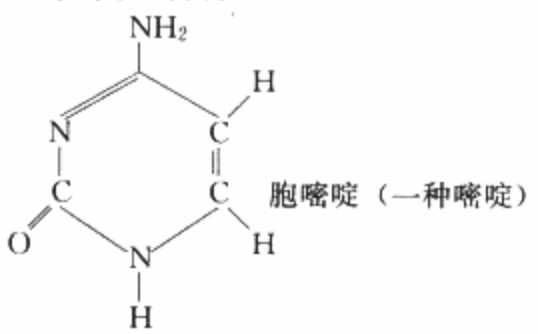


酸和胺结合形成酰胺。与胺不一样，酰胺在水中不带电荷。一个例子就是一个蛋白质中将氨基酸连接起来的肽键。



酸　　胺　　酰胺

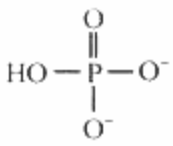
氮也出现在许多环状复合物中，包括重要的核酸组分：嘌呤和嘧啶。



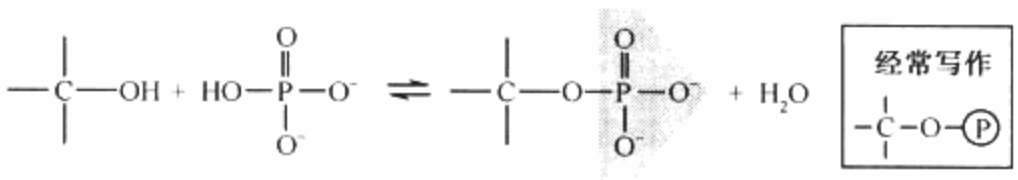
胞嘧啶（一种嘧啶）

磷酸

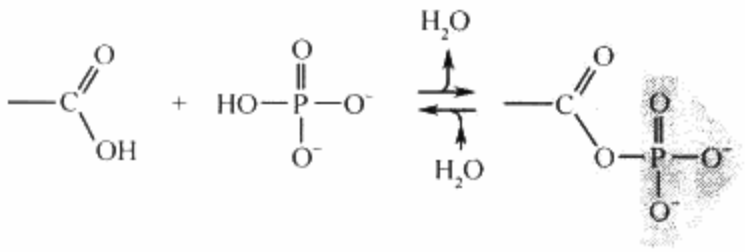
无机磷酸根是由磷酸 H_3PO_4 产生的稳定的离子。无机磷酸根通常写成 Pi。



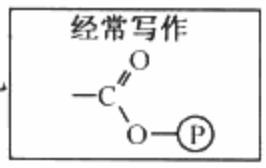
磷酸酯可以由 1 个磷酸根和 1 个自由羧基结合形成。磷酸根基团通常以这种方式结合到蛋白质上。

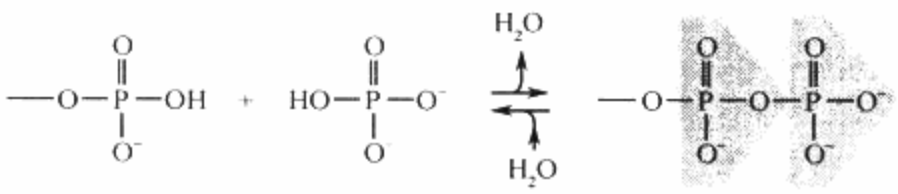


1 个磷酸基团结合 1 个羧基或 2 个或更多的磷酸基团就形成酸酐。

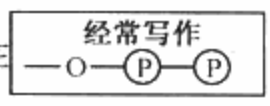


一些代谢物中存在高能乙酰磷酸键（羧基-磷酸酐）





一些分子如 RNA 中存在磷酸酐——一种高能键



原子的行为常常表明它的半径似乎是固定的

当在两个原子之间形成共价键时，共用电子将这些原子的核拉得非常近。但是在细胞中，这些原子迅速地相互推挤，大多数原子位于单独的分子中。当这样的两个原子接触时会发生什么呢？

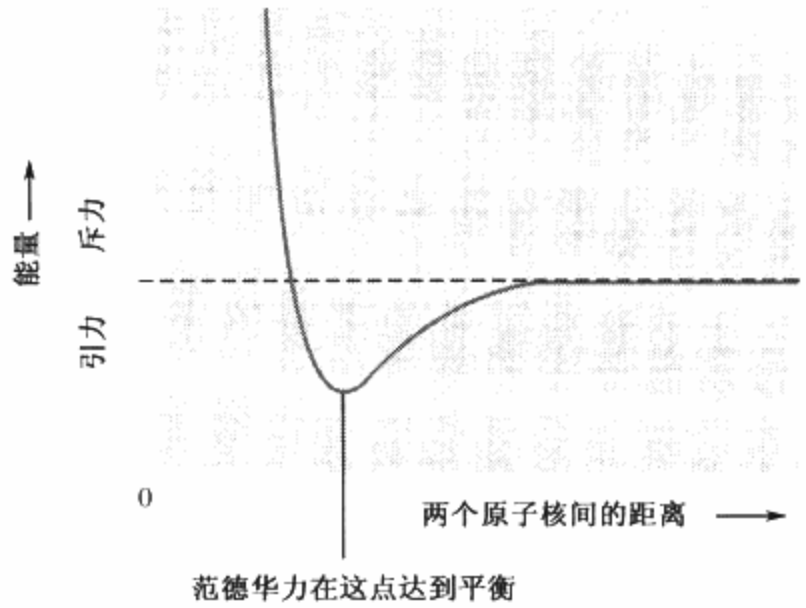


图 2-11　两个原子之间范德华力的平衡

随着两个原子的核相互靠近，由于它们波动的电荷，它们刚开始表现出弱的结合相互作用。然而，如果它们靠得太近，相同的原子就强烈地互相排斥。这些范德华引力和范德华斥力的平衡发生在图中所标的能量最低值处

为了简化和清晰，原子和分子通常以一种示意图来表示——结构式的直线图或球棒模型。然而，通过使用所谓的空间结构模型可以更为精确地表示原子和分子。在空间结构模型中，用实心球体表示电子云的半径，在半径处的强斥力阻止任意的第二个非键合原子靠近，即所谓的原子的范德华半径。范德华半径是可能存在的，因为 2 个原子互相接近时斥力强度即迅速地增加。在稍远处，任意 2 个原子之间具有弱引力，称为范德华引力。因此，存在一个斥力和引力精确平衡的距离，从而每个原子与第二个非键合元素的原子相互作用产生最低能量（图 2-11）。

根据我们的目的，在本书中我们以直线图、球棒模型或空间结构模型表示小分子。为了比较这三种表示方式，在图 2-12 中水分子以所有三种方式来表示。当表示非常大的分子（如蛋白质）时，我们经常需要进一步简化所用的表示法（如专题 3-2）。

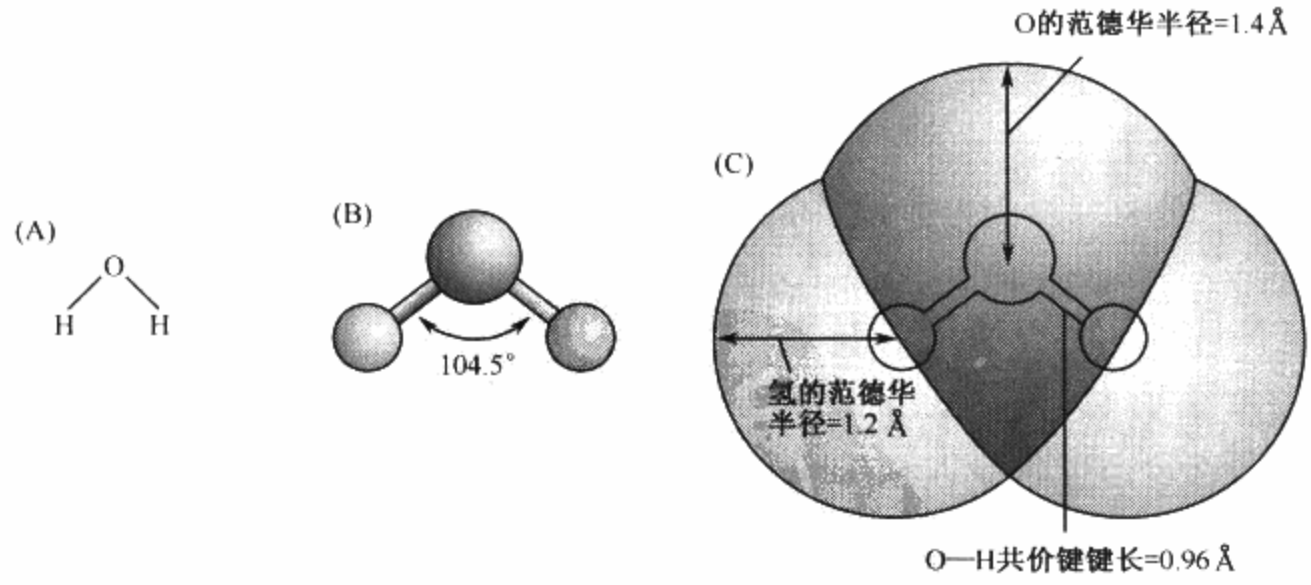


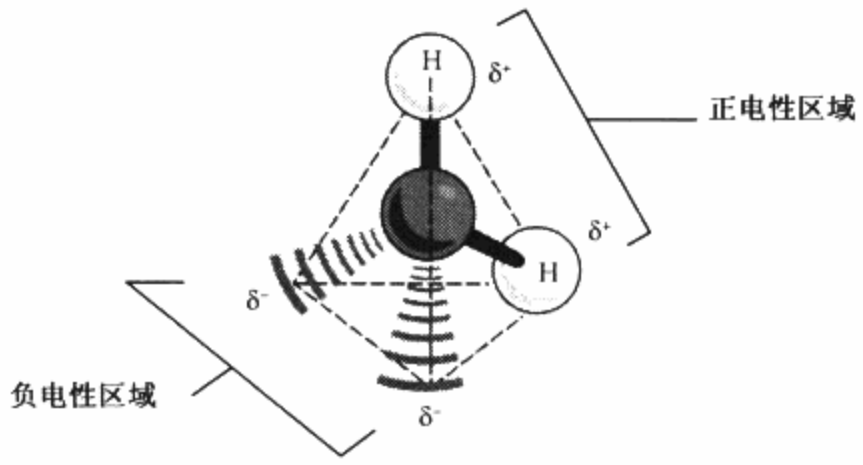
图 2-12　水分子的三种表示法

（A）结构式的直线图。每个原子用其标准符号表示，每条直线表示连接两个原子的一个共价键。（B）球棒模型。原子用任意直径的球表示，这些球是由表示共价键的棒连接的。和（A）不同，这种模型可以准确地表示键角（图 2-8）。（C）空间结构模型。该模型可以准确地表示键长和范德华半径

专题 2-2　水及其对生物分子行为的影响

水

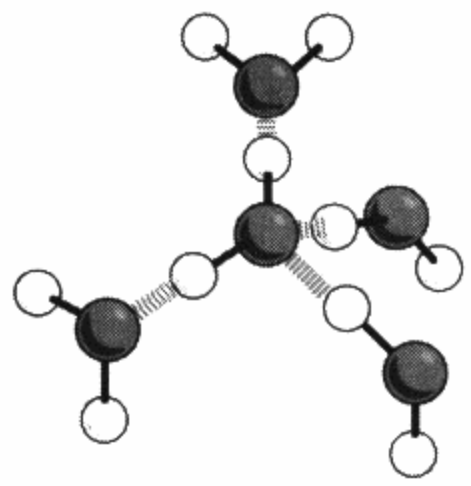
共价连接的 2 个原子，对于共价键的电子可能有不同的吸引力。在这些情况下，该共价键具有极性，一端带有少量的负电荷（S^-），另一端带有少量的正电荷（S^+）。



尽管水分子总体上不带电荷（电子数与质子数相等），但是由于电子不对称分布，所以导致分子具有极性。氧原子核将电子从氢原子核吸引开，使得氢原子核带少量的净正电荷。

水的结构

水分子瞬时连接在一起形成氢键晶格。甚至在 37℃，15%的水分子与其他的 4 分子水连接形成称为“闪动簇团”的短暂集合。

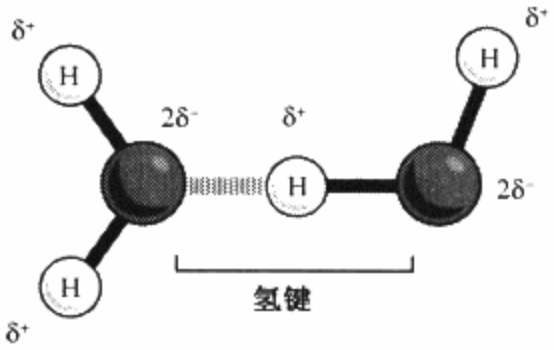


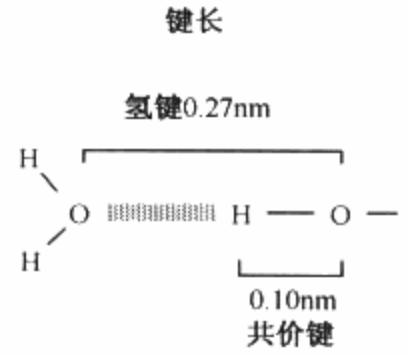
水的黏性导致水具有很多不同寻常的特性。例如，高的表面张力、特定的热和蒸发热。

氢键

因为 H_2O 是极性的，所以 2 个相邻的水分子能形成一种称为氢键的连接。氢键的强度只有共价键的 1/20 左右。

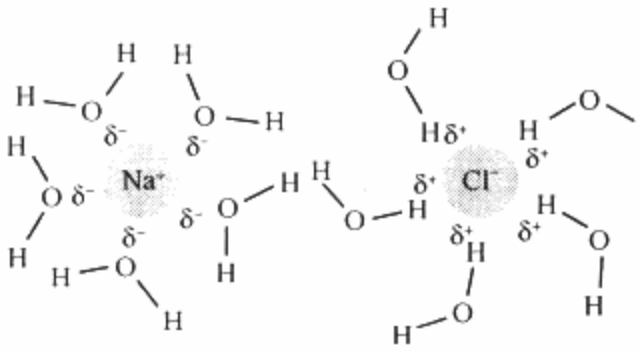
当 3 个原子位于一条直线时氢键最强。



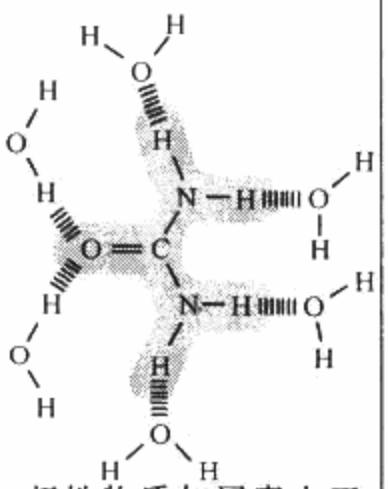


亲水分子

在水中容易溶解的物质称为亲水的。亲水分子是由通过电荷效应吸引水分子的离子或极性分子组成的。水分子在固体物质的表面围绕每个离子或极性分子，并将离子或分子带进溶液。



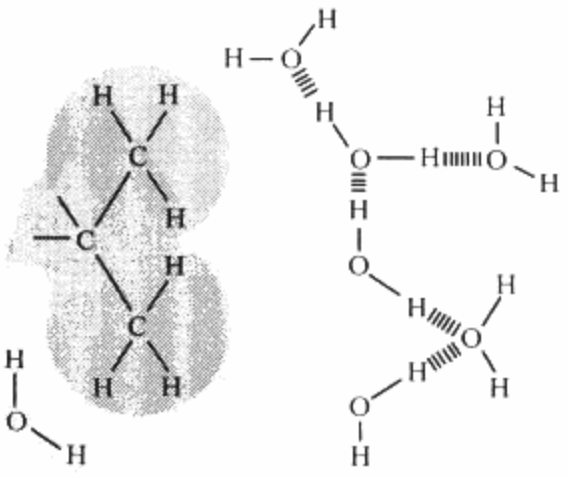
离子物质如氯化钠由于每个正（Na^+）或负（Cl^-）离子的电荷吸引水分子的缘故，所以可以溶解。



极性物质如尿素由于它们的分子能与周围水分子间形成氢键，所以可以溶解。

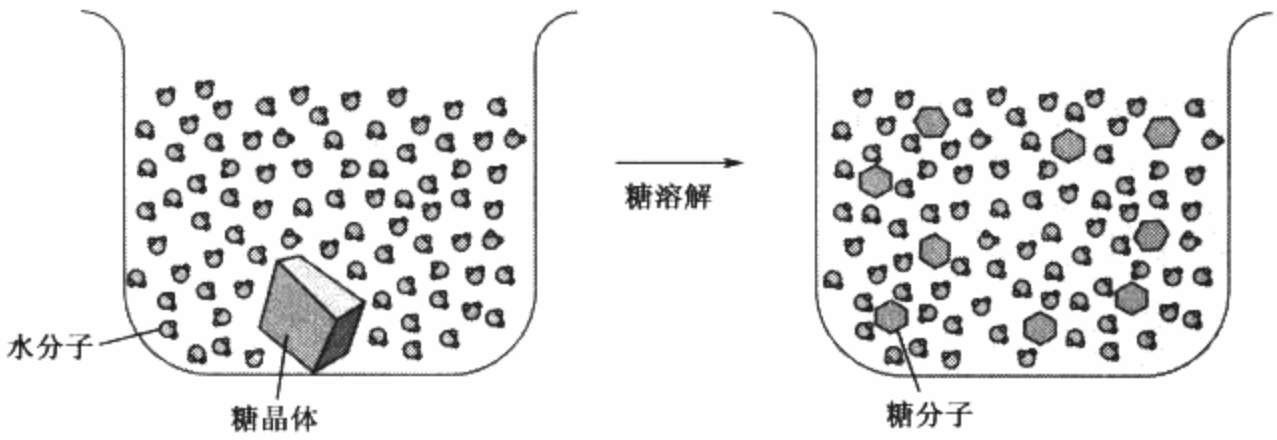
疏水分子

含有非极性键占多数的分子通常在水中是不可溶的，因此称为疏水的。这是真的，特别对于烃来说，烃含有许多 C—H 键。水分子不吸引烃分子，所以水分子不易于围绕烃分子并将它们带进溶液。



作为溶剂的水

许多物质都能溶于水，如糖。也就是说，它们的分子能够彼此分离，各自被水分子包裹起来。



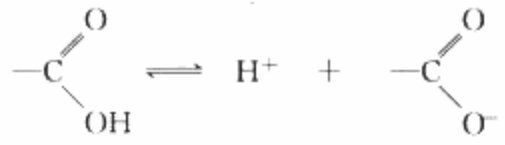
当物质溶解于液体中，混合物定义为溶液。溶解的物质（本例中是糖）是溶质，而起溶解作用的液体是溶剂。由于水的极性键的存在，水对于许多物质来讲是一种极好的溶剂。

酸

能释放氢离子到溶液中去的物质称为酸。

$$\underset{\text{强酸}}{HCl} \longrightarrow H^+ + Cl^-$$

细胞内重要的酸都仅仅部分解离，因此它们是弱酸，如羧基（—COOH）解离后释放一个氢离子到溶液中去。

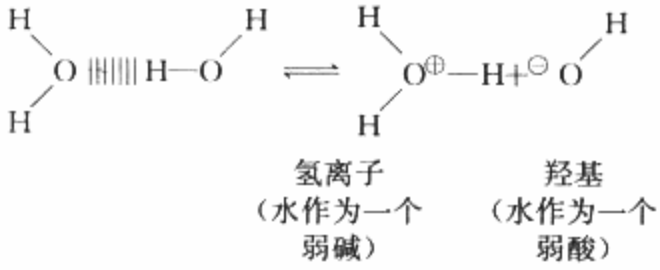


（弱酸）

注意这是一个可逆反应。

氢离子交换

带正电荷的氢离子能自发地从一个水分子转移到另一个水分子，因而生成两种不同的离子。

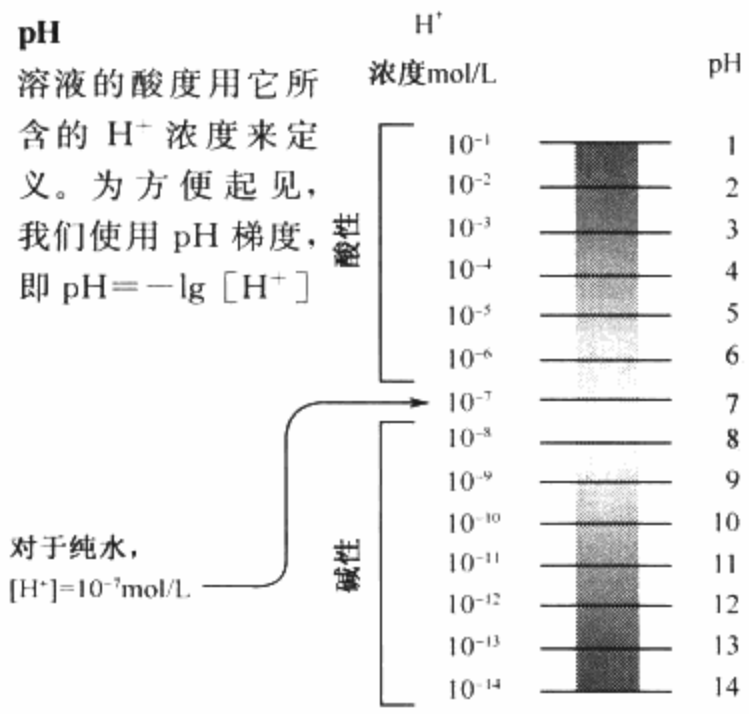


常写作：$H_2O \rightleftharpoons H^+ + OH^-$

由于这个过程迅速可逆，所以氢离子不停地穿梭于水分子之间。纯水含有稳态浓度的氢离子和氢氧离子（都是 10^{-7} mol/L）。

pH

溶液的酸度用它所含的 H^+ 浓度来定义。为方便起见，我们使用 pH 梯度，即 $pH = -\lg [H^+]$



碱

能减少溶液中氢离子数目的物质称为碱。某些碱，如氨能直接结合氢离子。

$$NH_3 + H^+ \longrightarrow NH_4^+$$

还有些碱，如氢氧化钠，通过产生 OH^- 与 H^+ 直接结合生成水，从而间接地减少溶液中的氢离子数目。

$$\underset{\text{强碱}}{NaOH} \longrightarrow Na^+ + OH^-$$

细胞内的许多碱部分解离，因而称为弱碱。例如，那些含有一个氨基基团（$-NH_2$）的化合物。这个氨基具有一种弱的趋势，能从水中可逆地接受一个 H^+，从而增加溶液中自由 OH^- 的数目。

$$-NH_2 + H^+ \rightleftharpoons -NH_3^+$$

水是细胞中含量最丰富的物质

水约占细胞重量的70%，并且大多数的细胞反应发生在液体环境中。地球上的生命起始于海洋，这一原始环境条件对生物的化学产生了永久的影响。因此，生命依赖于水的性质。

在每个水分子（H_2O）中，2个H原子通过共价键与O原子连接（图2-12）。2个键是高度极性的，因为O原子强烈吸引电子，而H对电子的吸引弱。因此，水分子中存在电子的不均匀分布，2个H原子带正电荷，而负电荷在O原子上的优势（图2-10）。当一个水分子的带正电荷区域（即其中的一个H原子）接近第二个水分子的带负荷区域（即O原子）时，它们之间的电荷吸引能产生一种称为氢键的弱键。与共价键相比，氢键要弱得多，容易被分子热能产生的随机热运动所打断。所以每个键的持续时间非常短。但许多弱键的加合效应远非微弱。每个水分子能通过它的两个H原子与另外2个水分子形成氢键，形成一个氢键不断被打断和生成的网络（专题2-2），正是因为将水分子连接到一起的氢键，水在室温下是液体，具有高沸点和高表面张力，而不是气体。

含有极性键并能与水形成氢键的分子（如乙醇）易溶于水。正如以前提过的，带正电荷或负电荷的分子（离子）一样与水互相作用。这些分子称为**亲水的**（hydrophilic），意思是指它们是喜好水的。细胞的液体环境中的大部分分子属于该范畴，包括DNA、RNA和大多数蛋白质。然而疏水（hydrophobic，憎水）分子是不带电荷的，形成少数或不形成氢键，因此不溶于水。碳氢键是一个重要的例子（专题2-1）。在这些分子中，氢原子通过一个非极性键与C原子共价连接。因为H原子几乎不带净的正电荷，它们不能与其他分子形成有效的氢键。这使得碳氢完全疏水，这一性质为细胞利用，细胞膜是由具有长的碳氢尾的分子构成的，这一点我们将在第10章介绍。

某些极性分子在水中形成酸和碱

当具有一个高度极性的共价键（氢原子和另一个原子之间）的大分子溶于水中时，发生一种最简单的化学反应，该化学反应对细胞非常重要。该分子中的氢原子将其电子给予相邻原子，所以以几乎裸露的带正电荷的氢核存在，即质子（proton，H^+）。当极性分子被水分子环绕时，相邻水分子的O原子上的部分负电荷吸引质子，质子能从它的原初配体上解离，从而与水分子上的氧原子结合，形成水合氢离子（hydronium ion，H_3O^+）［图2-13（A）］。逆反应也非常容易发生，所以人们必定会想到一种平衡状态，该平衡状态的数亿个质子不断地从溶液中的一个分子移到另一个分子。

溶于水时，能释放质子形成H_3O^+的物质称为酸（acid）。H_3O^+浓度越高，溶液的酸性越强。甚至在纯水中也存在H_3O^+，浓度为10^{-7} mol/L，这是质子从一个水分子转移到另一个分子的结果［图2-13（B）］。按照惯例，H_3O^+浓度通常是指H^+浓度，即使溶液中大多数H^+是以H_3O^+形式存在。为了避免使用不便于表示的数字，H^+浓度用其对数值表示，称为pH值，如专题2-2中所示。纯水的pH值为7.0。

因为水合氢离子中的质子能容易地转移到细胞内许多种分子，改变它们的特性，所以细胞内H_3O^+浓度（酸性）必须受到严密的调控。如果溶液中H_3O^+浓度低，能给出质子的分子更容易给出质子；如果溶液中H_3O^+浓度高时，能接受质子的分子易于接受质子。

与酸相对的是**碱**（base）。正如酸的界定性质是将质子给予水分子从而升高H_3O^+的浓度，碱的界定性质是升高氢氧根离子的浓度——氢氧根离子是通过从水分子去掉一个质子而形成。因为氢氧化钠在液体溶液中解离形成Na^+和OH^-，所以氢氧化钠（NaOH）是碱性的

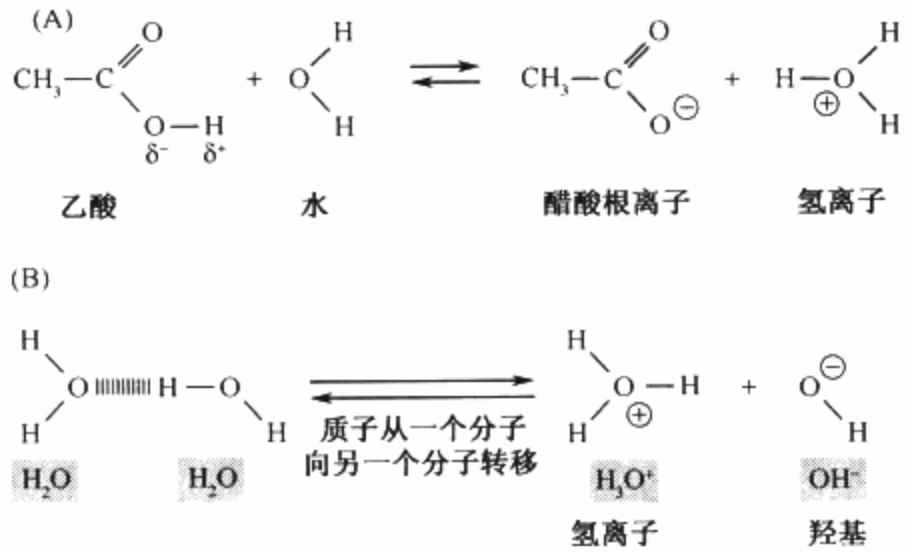


图 2-13　水中的酸

（A）当乙酸溶于水时发生的反应。（B）水分子不断地相互交换质子，以形成水合氢离子和氢氧根离子。这些离子再迅速重新结合以形成水分子

（也用术语 alkaline 表示）。另一类在活细胞中特别重要的碱含有 NH_2 基团。这些基团能通过从水中获得一个质子产生 OH^-：$-NH_2 + H_2O \longrightarrow -NH_3^+ + OH^-$。

因为一个 OH^- 与一个 H_3O^+ 结合形成两个水分子，所以 OH^- 浓度的升高促使 H_3O^+ 浓度的下降，反之亦然。纯的水溶液含有相等低浓度（10^{-7} mol/L）的两种离子；纯水既不是酸性的，也不是碱性的，因此是中性的，pH 为 7.0。细胞的内部被维持接近中性。

4 种非共价相互作用帮助细胞内分子结合

在溶液里，共价键比原子之间的吸引力强 10～100 倍，使共价键的连接能界定一个分子与另一个分子之间的界限。但是生物界大都依赖于不同分子之间相互的特异结合。这种特异结合是由一群非共价吸引力介导的，这些非共价吸引力本身非常弱，但它们的键能可以加和，在两个单分子之间产生一种有效的力。我们已经介绍了三种非共价吸引力：离子键、氢键和范德华力。表 2-2 比较了该三种键在水中和不在水中与共价键的键能。它们在生物系统中极其重要，因而我们在此总结它们的性质。

表 2-2　共价和非共价的化学键

键型	键长	键能	
		真空	水中
共价键	0.15	90	90
非共价键：离子键	0.25	80	3
氢键	0.30	4	1
范德华力（每原子）	0.35	0.1	0.1

离子键（ionic bond）　离子键是指带相反电荷的原子之间纯粹的静电吸引力。正如我们介绍的 NaCl 一样，离子键在非水溶液中是非常强。然而，极性的水分子聚集在带电荷的离子和具有永久偶极的极性分子周围（图 2-14），这大大降低了这些离子之间潜在的相互吸引力（表 2-2）。

氢键（hydrogen bond）　典型氢键的结构如图 2-15 所示。氢键是一种特殊的极性相互作用，两个带负电荷的原子部分共用一个带正电荷的氢原子。氢键上的氢可认为是从供体原子上部分解离出来的质子，使第二个受体原子能共用该质子。与典型的静电相互作用不一样，氢键具有很强的方向性，当成键的三个原子在一条直线时氢键最强。正如已讨论的，水通过与有关分子形成竞争性的氢键相互作用可以减弱这些氢键（表 2-2）。

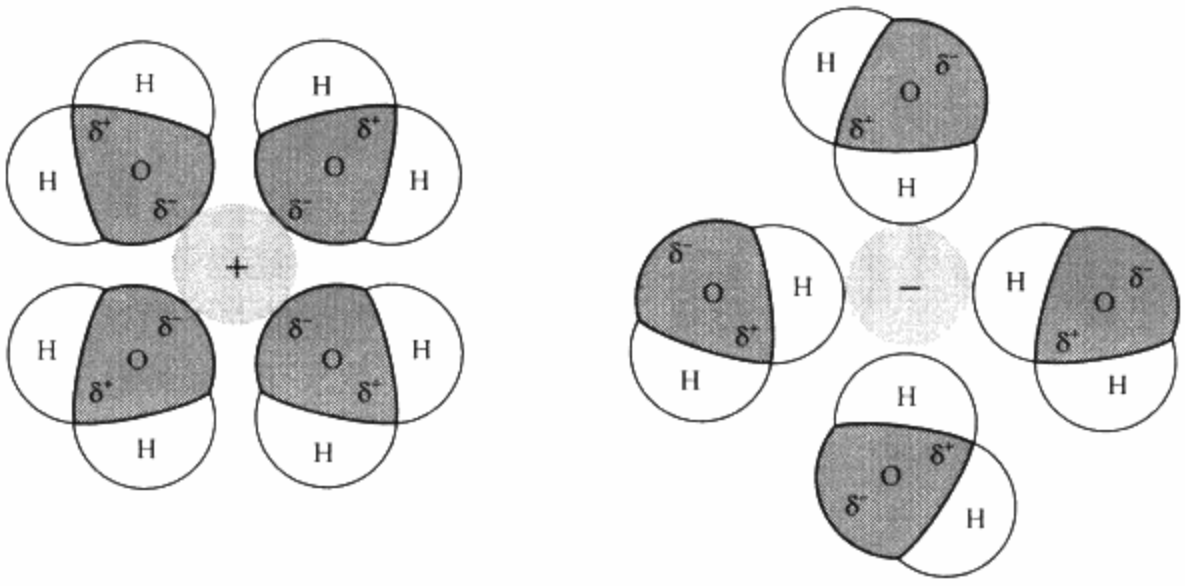
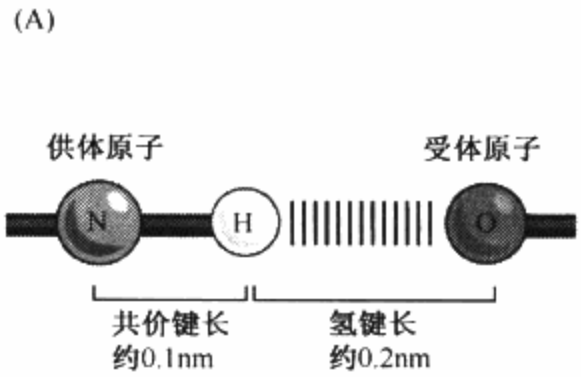


图 2-14　水分子上的偶极如何定向以降低带相反电荷的离子或极性基团之间的亲和力



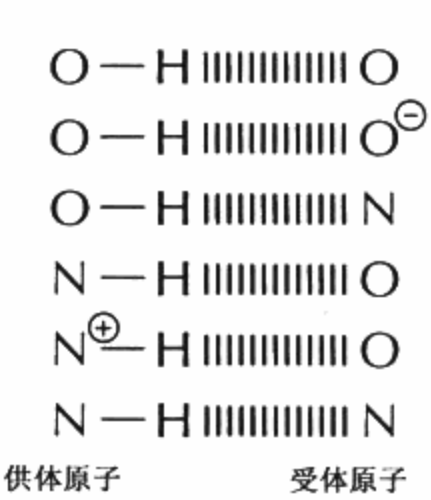


图 2-15　氢键

（A）一个典型氢键的球棒模型。图中氢原子和氧原子之间的距离小于它们的范德华半径的总和，表明它们部分共用电子。（B）细胞内最常见的氢键

范德华力（van der Waals attraction）　任何非极性原子的电子云可以变动，产生瞬时偶极。该瞬时偶极可以瞬时诱导附近的原子产生相反极性的瞬时偶极。这种相互作用产生原子之间非常弱的吸引力。但是，当两个表面非常吻合时许多原子能同时相接触，所以净结果是非常显著的。这些所谓的范德华力并不被水减弱（表 2-2）。

疏水作用（hydrophobic force）　是在水中将分子聚集在一起中起重要作用的第四种作用。这种作用是由非极性表面被挤出氢键的水网络所引起的，非极性表面物理上干扰了水分子之间高度有利的相互作用。因为将两个非极性表面聚集到一起减少了它们与水的接触，所以疏水作用是相当非特异性的。但是，在第 3 章我们将看到疏水作用对于蛋白质分子的正确折叠是很关键的。

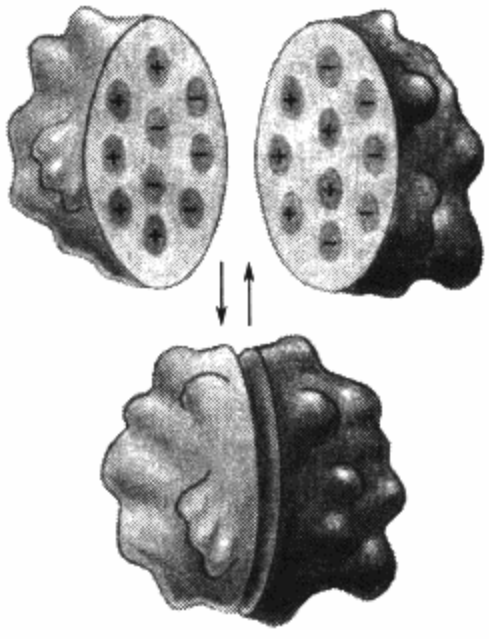


图 2-16　具有互补表面的两个大分子如何通过非共价的相互作用紧密结合

在该图解中，正和负被用来标注当分子配对时能形成的吸引相互作用

专题 2-3 概括了刚介绍的 4 种相互作用。图 2-16 以图解的方式解释了许多这些相互作用加起来如何使两个大分子的互补表面结合，即使每种作用本身太弱以至于对互补表面的结合没有影响。

细胞是由碳化合物构成

了解了原子结合成小分子的方式以及这些分子在溶液中的作用方式，我们现在介绍细胞里的主要小分子及其生物学作用。我们将看到几种基本类型分子（由几种不同元素构成）产生出所有生物所表现出的非常丰富的形式和行为。

如果我们不考虑水的话，细胞中几乎所有的分子都是以碳为基础。在所有的元素中碳是突出的元素，因为它能形成大分子；硅其次。因为碳小，且其最外电子层带有 4 个电子和 4 个空缺，碳原子能与其他原子形成 4 个共价键。最重要的是，一个碳原子可通过高度稳定的 C—C 共价键与另一个碳原子连接从而形成链和环，因此可产生大而复杂的分子，分子的大小没有明显的上限（专题 2-1）。细胞产生的碳化合物（无论大小）称为有机分子。

某些原子的组合，如甲基（$—CH_3$）、羟基（—OH）、羧基（—COOH）、酮基（—C═O）、磷酸基（$—PO_3^{3-}$）和氨基（$—NH_2$），重复出现于有机分子中。每种**化学基团**（chemical group）具有不同的化学性质和物理性质，这些性质影响该基团所在分子的行为。专题 2-11 总结了最常见的化学基团以及它们的某些性质。

细胞含有 4 大类主要家族的有机小分子

细胞的有机小分子是以碳为基础的化合物，相对分子质量范围为 100～1000，最多含有大约 30 个碳原子。它们自由地存在于溶液中，具有许多不同的命运。某些被用作单体亚基以构建细胞的巨大的多聚体大分子——蛋白质、核酸和大的多糖。其他充当能源，在胞内代谢途径的迷宫中被分解和转化成其他的小分子。许多小分子在细胞中具有一种以上的功能，如充当大分子的潜在亚基和能源。有机小分子的含量远远低于有机大分子，大约只占细胞内有机物质总量的 1/10（表 2-3）。据粗略估计，一个典型的细胞内可能有 1000 种不同的小分子。

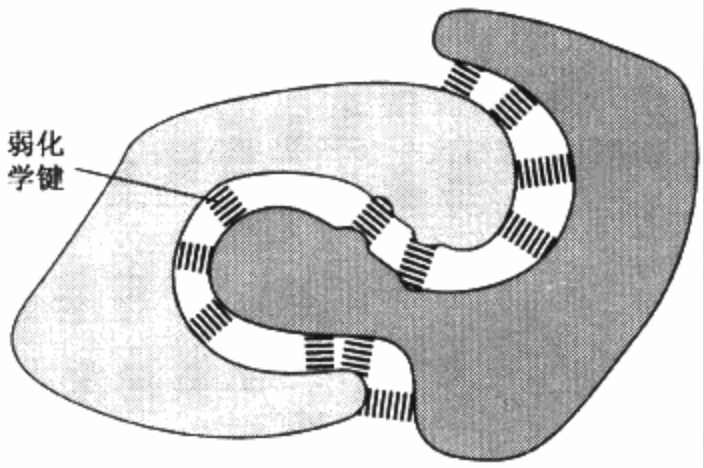
表 2-3　细菌细胞的大概的化学组分

	占总细胞重量的百分比	每分子的种类数
水	70	1
无机离子	1	20
糖及其前体	1	250
氨基酸及其前体	0.4	100
核苷酸及其前体	0.4	100
脂肪酸及其前体	1	50
其他小分子	0.2	～300
大分子(蛋白质、核酸和多糖)	26	～3000

专题 2-3　结合大分子的主要类型的弱非共价键

弱化学键

有机分子可以通过短距离的非共价作用力与其他的分子相互作用。

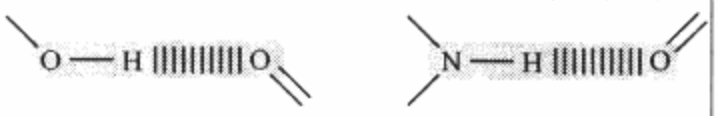


弱化学键的作用力大小不到强共价键的 1/20。只有当弱化学键同时形成时，才足够强烈以提供牢固的结合力。

氢键

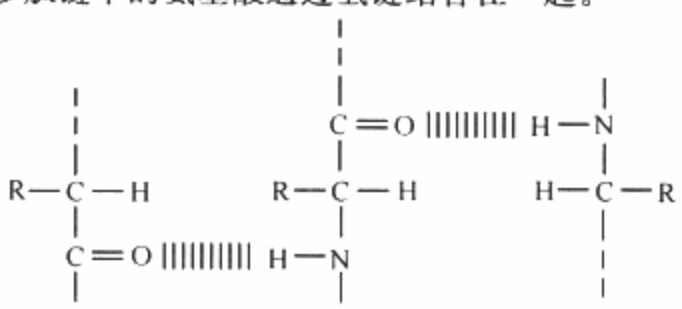
正如描述水时已提到的那样（专题 2-2），当两个吸引电子的原子（经常是氧或氮）将一个氢原子像三明治一样夹在中间时，就形成了氢键。

当 3 个原子位于一条直线时形成的氢键最强。

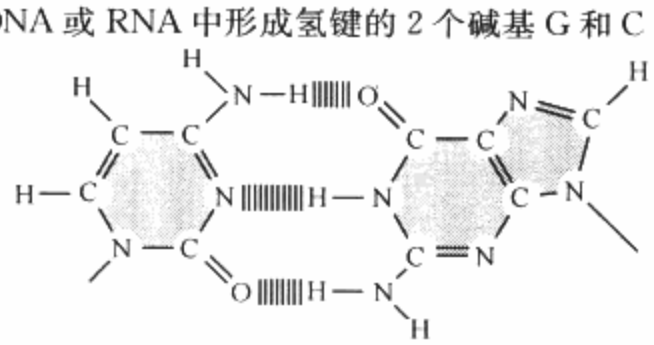


大分子中的实例：

多肽链中的氨基酸通过氢键结合在一起。

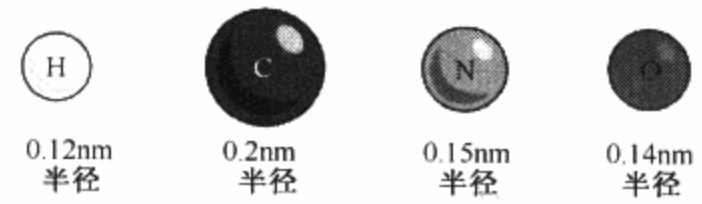


DNA 或 RNA 中形成氢键的 2 个碱基 G 和 C 。



范德华吸引力

两个原子彼此靠得太近，它们就会强烈地排斥对方。因为如此，一个原子通常可以看成有固定半径的圆球。每个原子特定的半径由独特的范德华半径规定。任何两个非共价结合的原子间的接触距离就是它们的范德华半径之和。

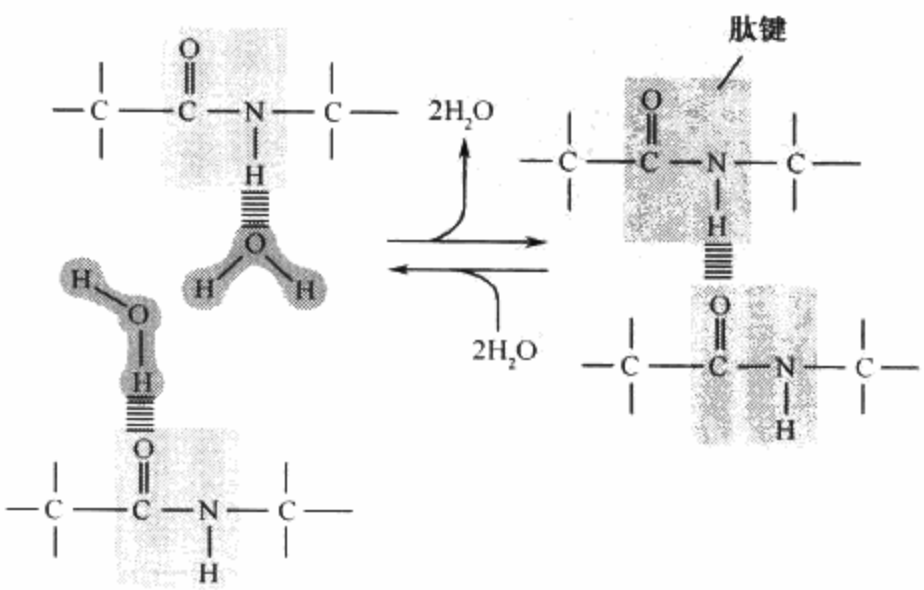


任何两个原子距离很近时，由于它们所带电荷不断发生改变，所以它们都表现出一种较弱的成键相互作用。这两个原子将以这种方式彼此吸引对方，直到它们的原子核间距接近它们的范德华半径之和为止。尽管单个看来，范德华吸引力非常微弱，当两个大分子表面非常吻合时，它就变得很重要了，因为将涉及很多的原子。

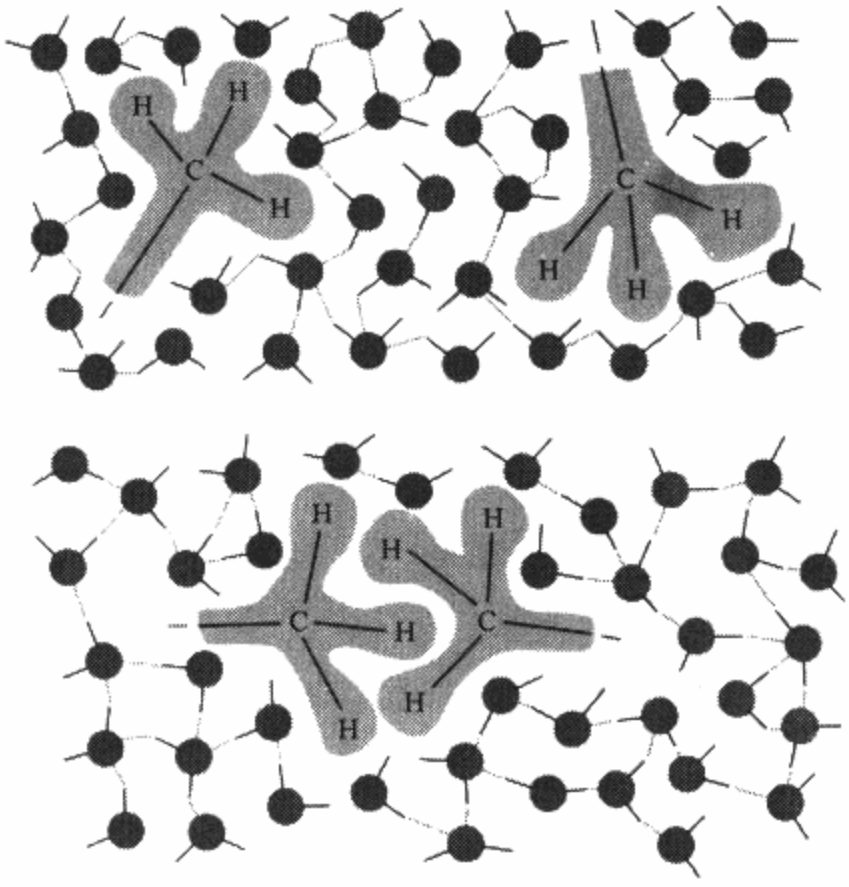
注意当两个原子之间形成共价键时，它们的中心（原子核）彼此靠得很近，远远小于两个原子的范德华半径之和。因此，

水中的氢键

任何分子如果能够彼此之间形成氢键，就同样能够与水分子之间形成氢键。由于和水分子之间的这种竞争，溶于水的两个分子之间形成的氢键就会相对较弱。



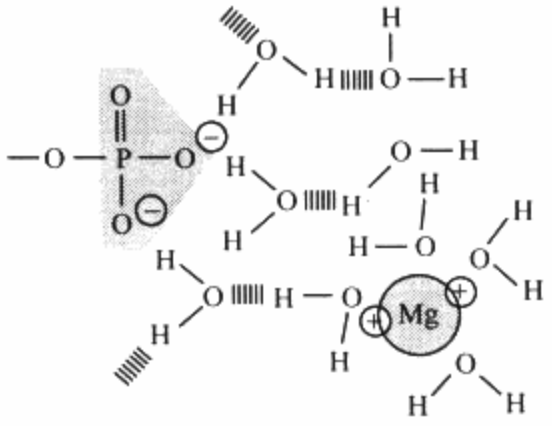
疏水作用



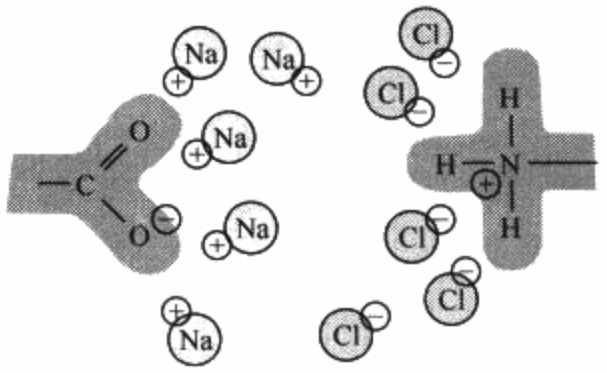
水迫使疏水基团集中在一起，因为这样可以减少这些基团对水中的氢键网的影响。疏水基团以这种方式聚集在一起，有时也称作被疏水键束缚在一起，即使由于水分子的排斥确实产生了相互吸引作用。

水溶液中的离子键

带电基团被它们与水分子间的相互作用保护起来。因此，水中离子键相当微弱。



同样，溶液中其他的离子可以在带电基团周围集聚，因而进一步减弱离子键。

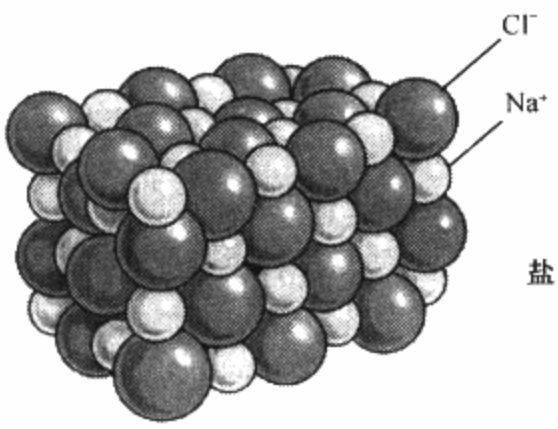


离子键

离子的相互作用发生在完全带电荷的基团之间（离子键）或者部分带电荷的基团之间。

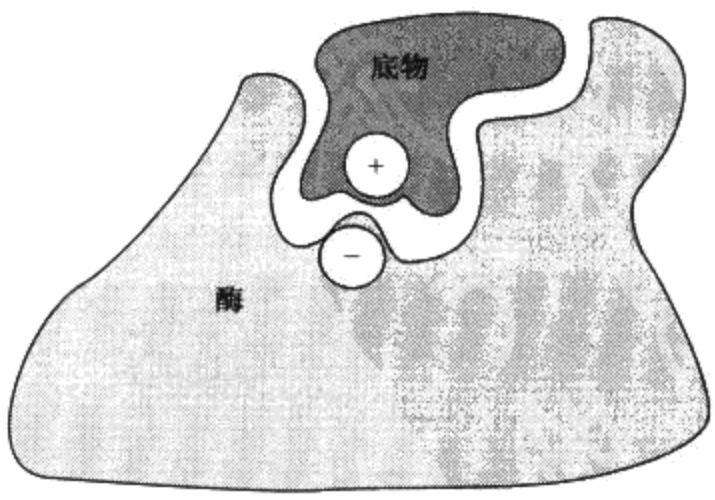
两种电荷 δ^+、δ^- 之间的吸引力随电荷间的距离增加而迅速减少。

在不存在水时，离子作用力非常大，因为如此，才会形成大理石、玛瑙这样的矿石。



盐的晶体

尽管离子键因水和盐的作用不断减弱，但是它在生物系统中仍相当重要；一个结合带正电荷底物的酶通常在适当的位置具有带负电荷的氨基酸侧链。



所有的有机分子是由一套相同简单化合物合成的，也能分解成相同的一套简单的化合物。有机分子的合成和分解是通过一系列的化学变化发生的。这些化学变化受到空间限制，并遵循明确的规律。因此，一个细胞内的化合物在化学上是相关的，大多数化合物能被归类为少数的不同家族。一般来说，细胞含有 4 大家族的有机小分子：糖、脂肪酸、氨基酸以及核苷酸（图 2-17）。尽管存在于细胞内的许多化合物并不属于这些范畴，这 4 家族的有机小分子和将它们连接成长链所形成的大分子一起占细胞的大部分质量（表 2-3）。

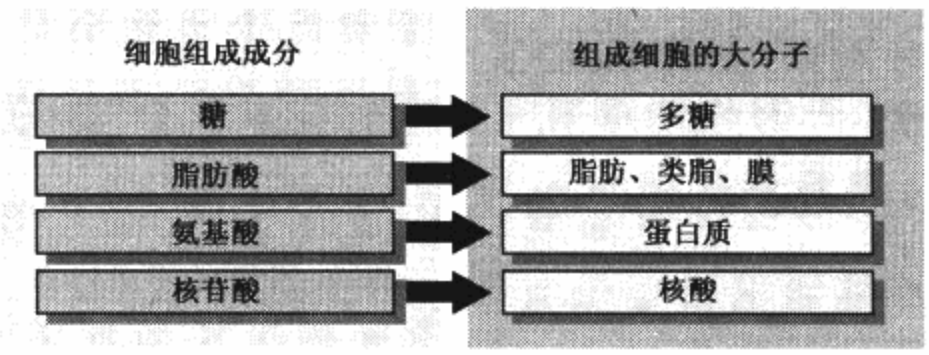


图 2-17　细胞内 4 大类有机小分子

这些小分子是单体构建元件或者亚单位，它们组成细胞内大多数大分子和其他的结构。其中有一些，如糖和脂肪酸，也是能量的来源

糖类物质为细胞提供能量来源，也是多糖的亚单位

最简单的**糖类物质**——单糖是一类化合物，可以用通用的公式 $(CH_2O)_n$ 表示，这里 n

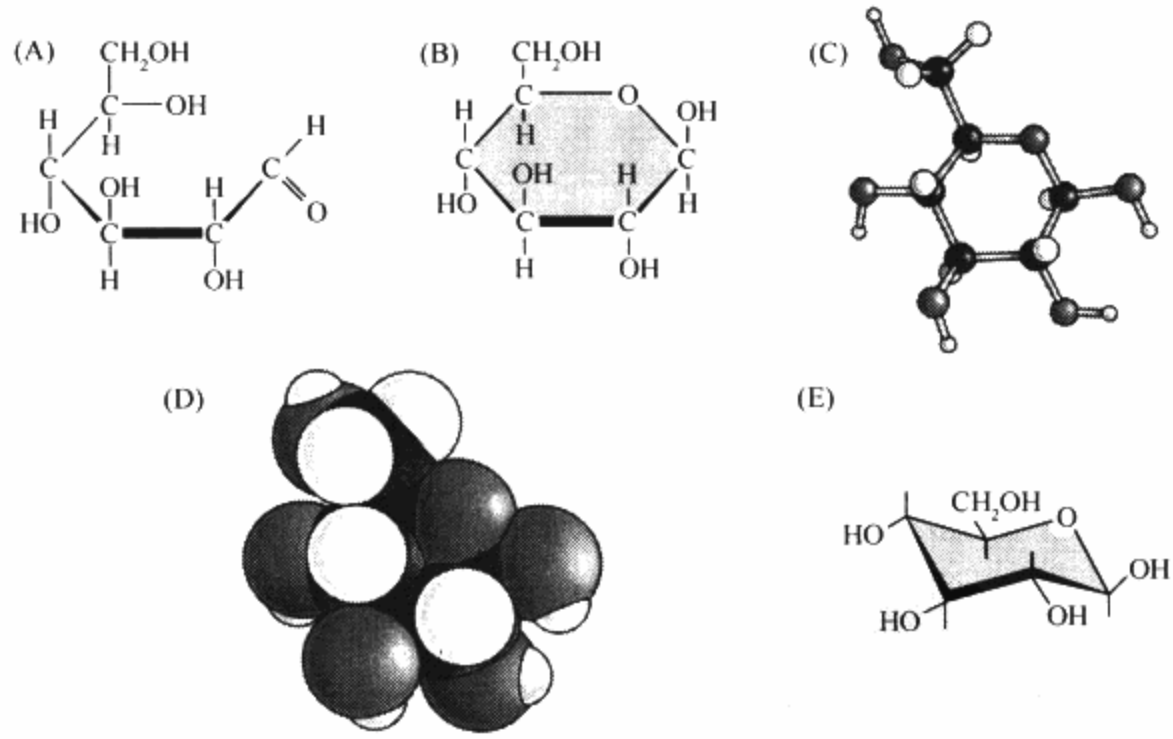


图 2-18　一种简单的单糖——葡萄糖的结构

正如前面图示水的结构一样（图 2-12），任何分子都可以用许多方式表示。在（A）、（B）和（E）所示的结构式中，用代表化学键的直线连接起来的化学符号表示原子。粗线在这里用来表示糖环平面，以强调—H 和—OH基团并非和环本身一样位于同一平面。（A）葡萄糖的开链形式，它与（B）中的更加稳定的环状形式处于平衡之中。（C）一种球棒模型，这里显示了原子在空间中的三维排布。（D）一种空间填充模型，它既描述了原子的三维排布情况，也用范德华半径表示出分子表面的轮廓。（E）椅式构型也可以用来表示环状分子，它比（B）中的结构式更加准确地反映出分子的几何结构。（C）和（D）中根据传统代表原子的颜色来描绘原子，如白色代表 H，黑色代表 C，红色代表 O，蓝色代表 N（图 2-8）

通常是 3、4、5、6、7 或 8。由于这个简单的通式，糖类以及它们组成的分子也称作碳水化合物。例如，葡萄糖的分子式是 $C_6H_{12}O_6$（图 2-18）。但是这个通式并没有完整定义该分子。因为同样的碳、氢、氧可以不同的方式通过共价键连接起来，产生不同形状的分子结构。如专题 2-4 所示，例如，葡萄糖仅仅通过改变特定—OH 基团可以转变成不同的糖——甘露糖或半乳糖，而且这些糖中的每一种都能以两种形式中的任何一种存在，分别称为 D 型和 L 型，它们互为对方的对映体。具有相同的化学结构式却有不同结构的一组分子称为异构体。一组对映体分子中属于异构体的分子称为旋光异构体。总体上讲，异构体在有机分子中广泛存在，它们对生成各种各样的糖类物质起主要作用。

专题 2-4 所示为糖类结构和化学的概括，糖类可以环状或者开链状结构存在。在它们的开链形式中，糖类含有许多羟基以及一个醛基$\left(H\diagdown C{=}O\right)$或者一个酮基$\left(\diagdown C{=}O\right)$。醛基或酮基发挥着特别的作用。首先，它可以与同一分子上的羟基发生反应，将分子转变为环式结构，位于环状结构中的最初醛基或酮基的碳可以看作唯一一个结合 2 个氧的碳。其次，一旦形式环状结构，相同的碳可与其他糖分子上拥有一个羟基的某个碳进一步连接起来，生成一个双糖。例如，蔗糖由一分子葡萄糖和一分子果糖组成。更大的糖多聚体包括寡糖（三糖、四糖等）和巨大的多糖，它由无数的单糖单位组成。

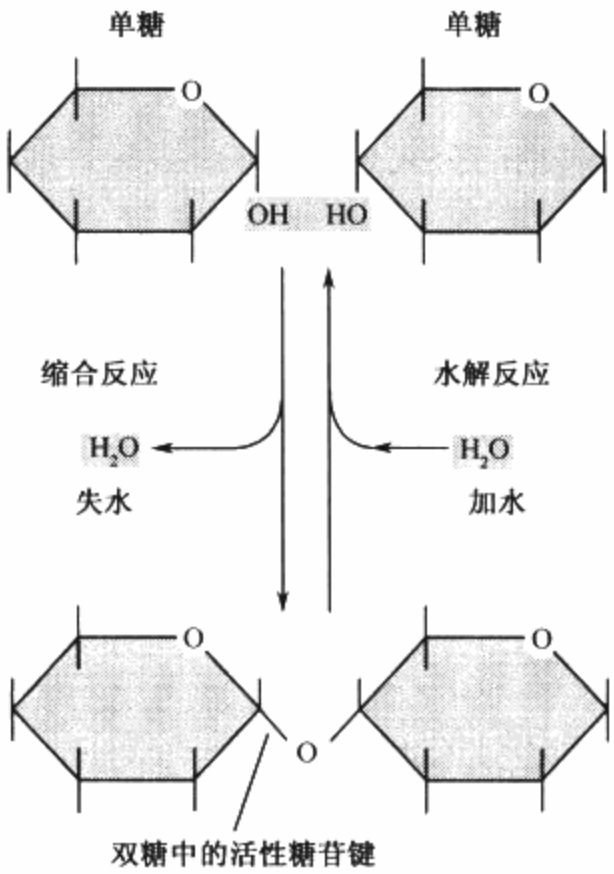


图 2-19　两个单糖形成一个二糖的反应

这个反应属于一种普遍类型的反应，称为缩合反应，在此反应中，2 个分子由于失去 1 分子水而连接起来。逆反应（反应中加水）称为水解反应。注意其中的 1 个碳（这里指左边这个）结合 2 分子氧，形成糖环（图 2-18）。如图中所示的，2 个糖分子之间形成的这种普通类型的共价键称为糖苷键（图 2-20）

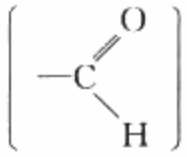
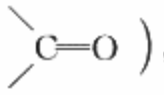
糖类物质彼此连接起来生成多聚体的方式证明生物化学键形成时存在某些一般特征。一个糖分子上的—OH 基团和另一个糖分子上的—OH 基团通过**缩合反应**，生成一个化学键。在此反应过程中，当化学键生成时，释放出一分子水（图 2-19）。其他生物多聚体，例如，核酸和蛋白质中的亚单位，也是通过缩合反应，并释放水分子，彼此连接起来的，所有这些缩合反应形成的化学键可以通过相反的过程，**水解反应**，被打断，在此过程中消耗水（图 2-19）。

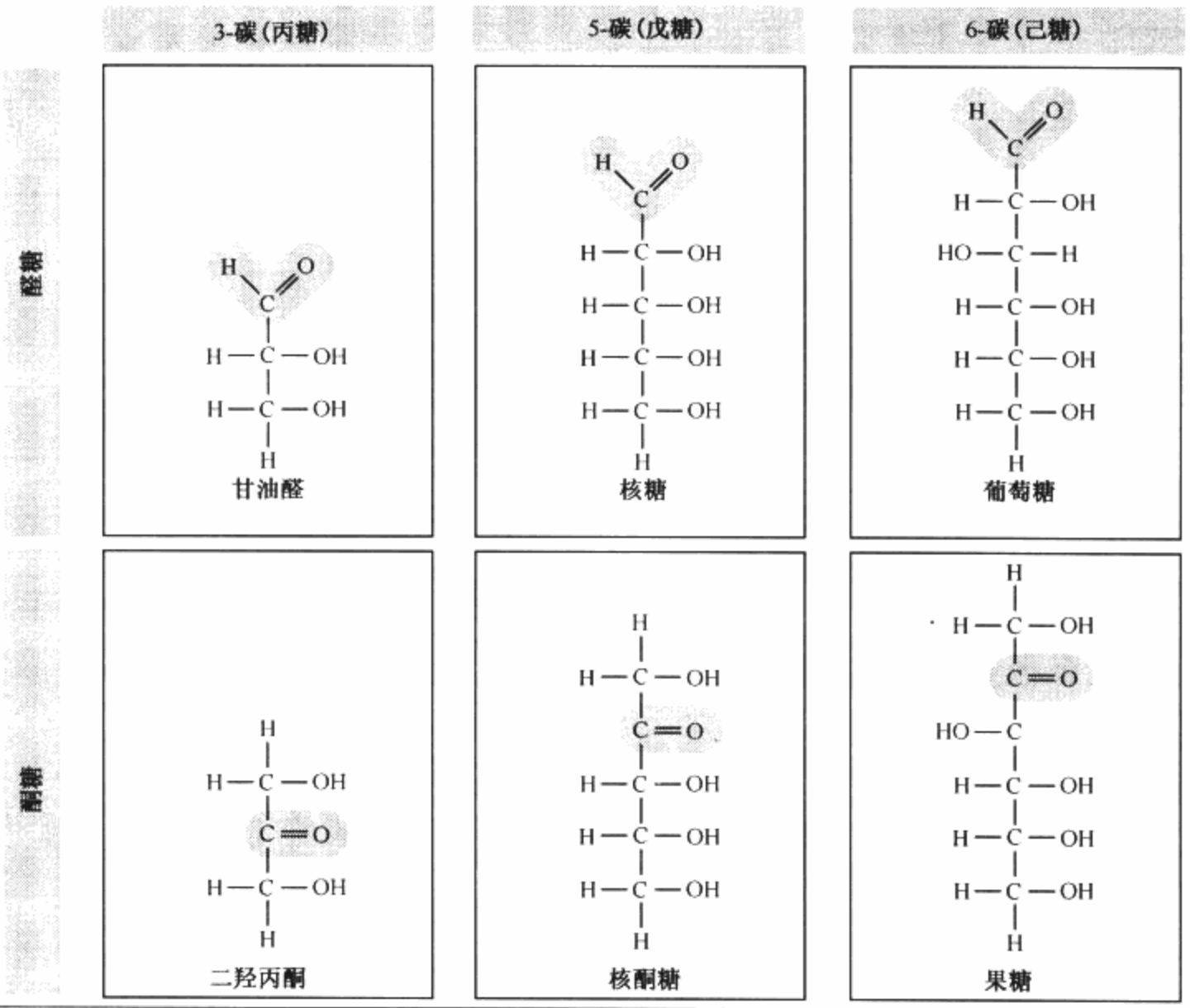
由于每个单糖拥有许多自由羟基，它们可以与其他的单糖（或者许多其他复合物）连接起来，所以糖聚合物能产生分支，而且可能出现的多糖结构也是多种多样的。即便是由 2 个葡萄糖残基也能形成 11 种不同的双糖（图 2-20）中，而三种不同的六碳糖（$C_6H_{12}O_6$），可以连接起来生成几千种三糖。因为如此，确定多糖中单糖单位的组成情况比确定 DNA 中的核苷酸序列要困难得多，后者中每个单体彼此以相同的方式连接起来。

葡萄糖单糖作为细胞能量的来源，起着重要作用，经过一系列反应，它断裂成更小的分子，释放出能量，细胞可以利用来作有用功，这些我们将在后面讨论。细胞利用仅由葡萄糖单位组成的简单多糖作为长期的能量储存形式，在动物中主要是糖原，在植物中主要是淀粉。

专题 2-4　细胞内常见糖的一些类型的概述

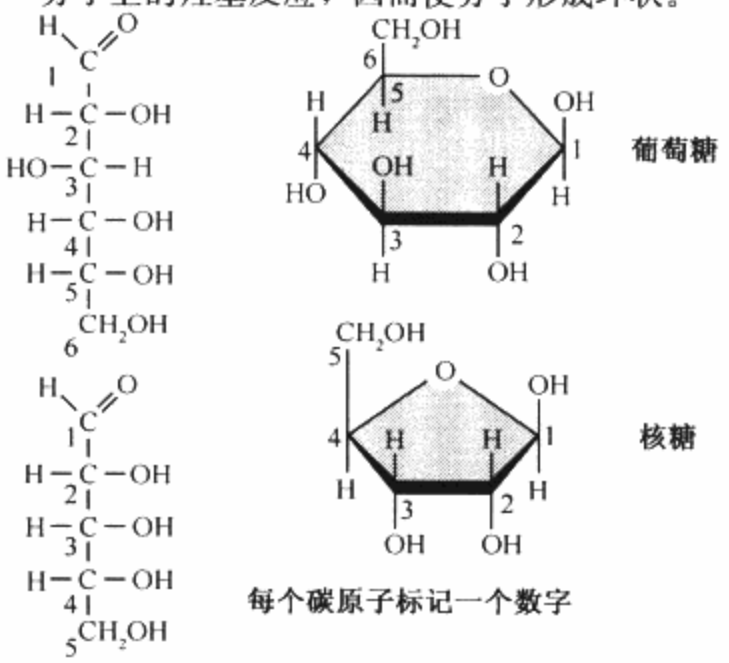
单糖

单糖通常具有一个通式$(CH_2O)_n$，这里 n 可以是 3、4、5、6、7 或 8，单糖具有两个或更多的羟基。单糖或者含有一个醛基  ，因而称为醛糖，或者含有一个酮基  ，因而称为酮糖。



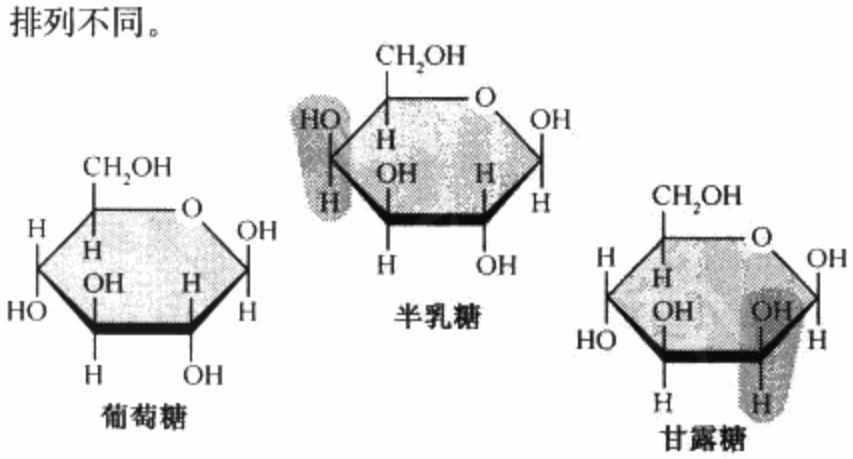
环的形成

在水溶液中，糖分子上的醛基或酮基倾向于与同一分子上的羟基反应，因而使分子形成环状。



异构体

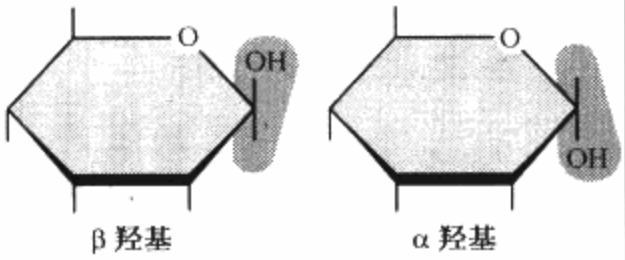
许多单糖仅仅在原子的空间排列上不同，也就是说，它们是异构体。例如，葡萄糖、半乳糖和甘露糖具有相同的公式（$C_6H_{12}O_6$），但是某一个或两个碳原子上的基团排列不同。



这些小差别仅仅使糖的化学性质发生较小的改变，但会使它们被酶和其他蛋白质识别，因此具有重要的生物作用。

α 和 β 连接

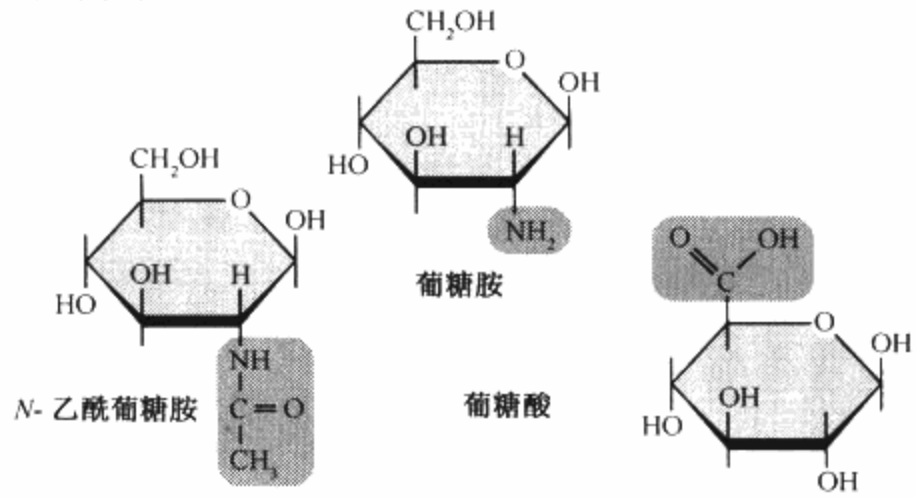
携带醛基或酮基的碳原子上的羟基可以迅速从一个位置变换到另一个位置。这两种位置称为 α 和 β。



一旦一个糖与另一个糖连接起来，α 或 β 形式也就确定不变。

糖衍生物

一个简单单糖的羟基可以被其他的基团取代。例如，



二糖

携带醛基或酮基的碳原子可以与另一个糖分子上的任一个羟基反应形成一个二糖。其中的连接键称为糖苷键。

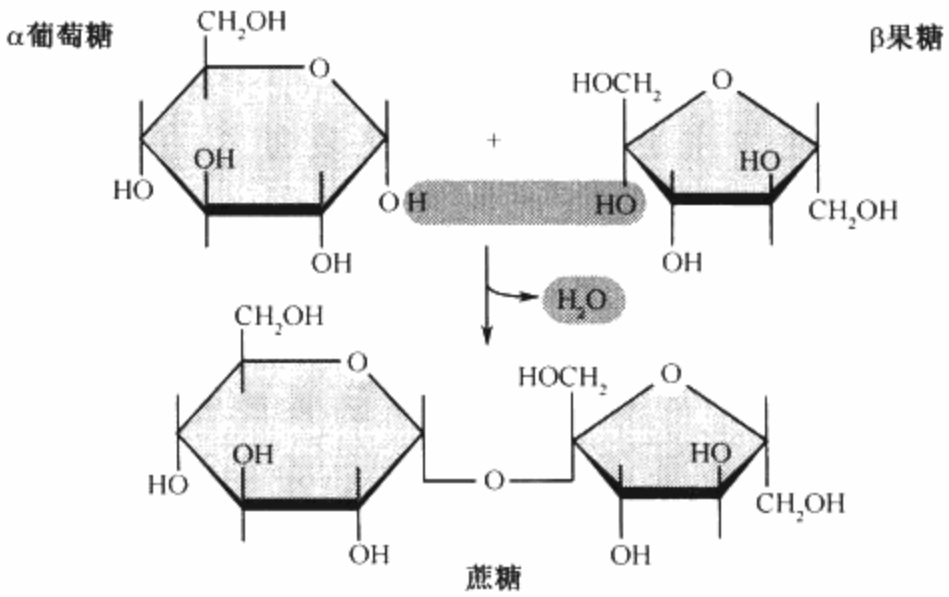
三种常见的二糖是

麦芽糖（葡萄糖＋葡萄糖）

乳糖（半乳糖＋葡萄糖）

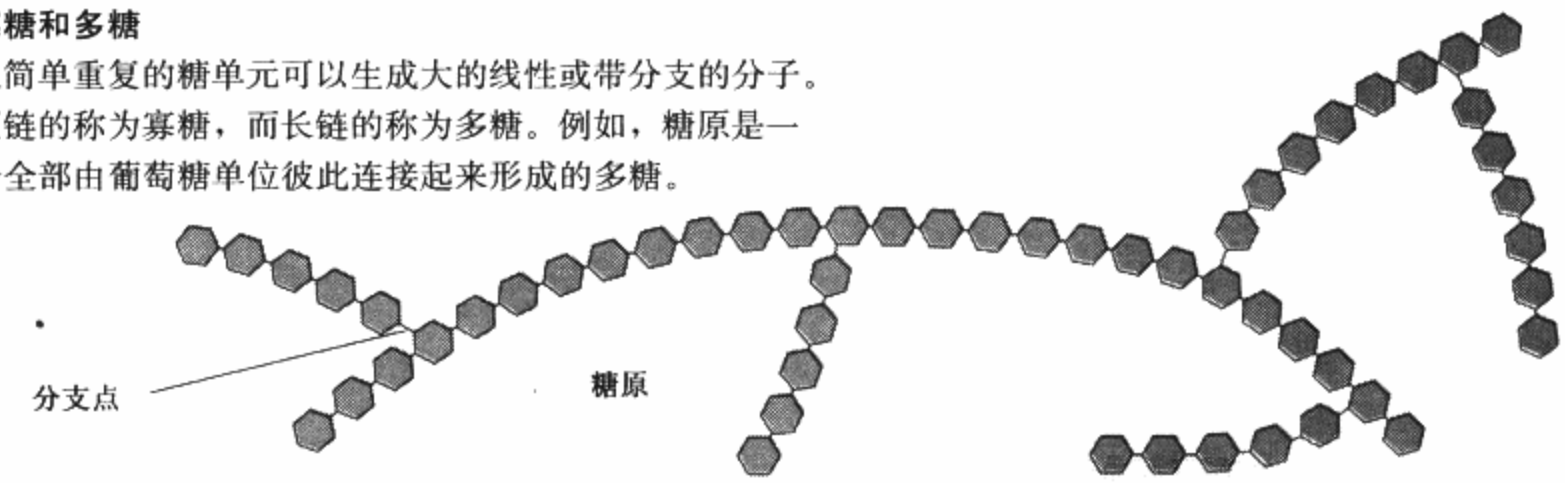
蔗糖（葡萄糖＋果糖）

形成蔗糖的反应如图所示。



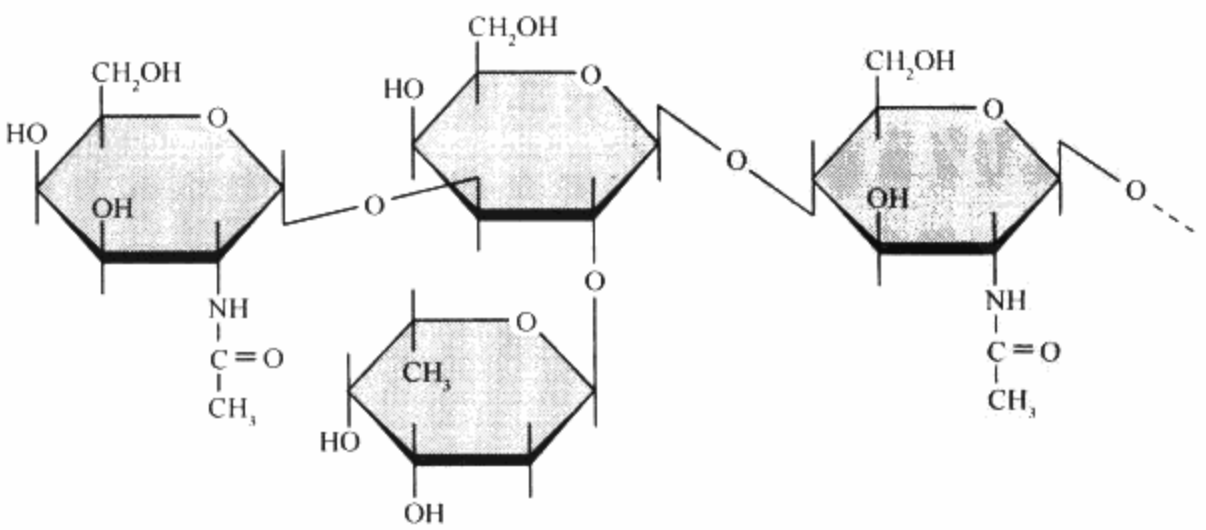
寡糖和多糖

从简单重复的糖单元可以生成大的线性或带分支的分子。短链的称为寡糖，而长链的称为多糖。例如，糖原是一个全部由葡萄糖单位彼此连接起来形成的多糖。



复杂的多糖

在许多情况下糖的序列不是重复的。许多不同的分子可能出现这种情况。这种复杂的寡糖经常与蛋白质或脂类结合，如图中所示的这个寡糖，它是细胞表面决定特定血型的分子的一部分。



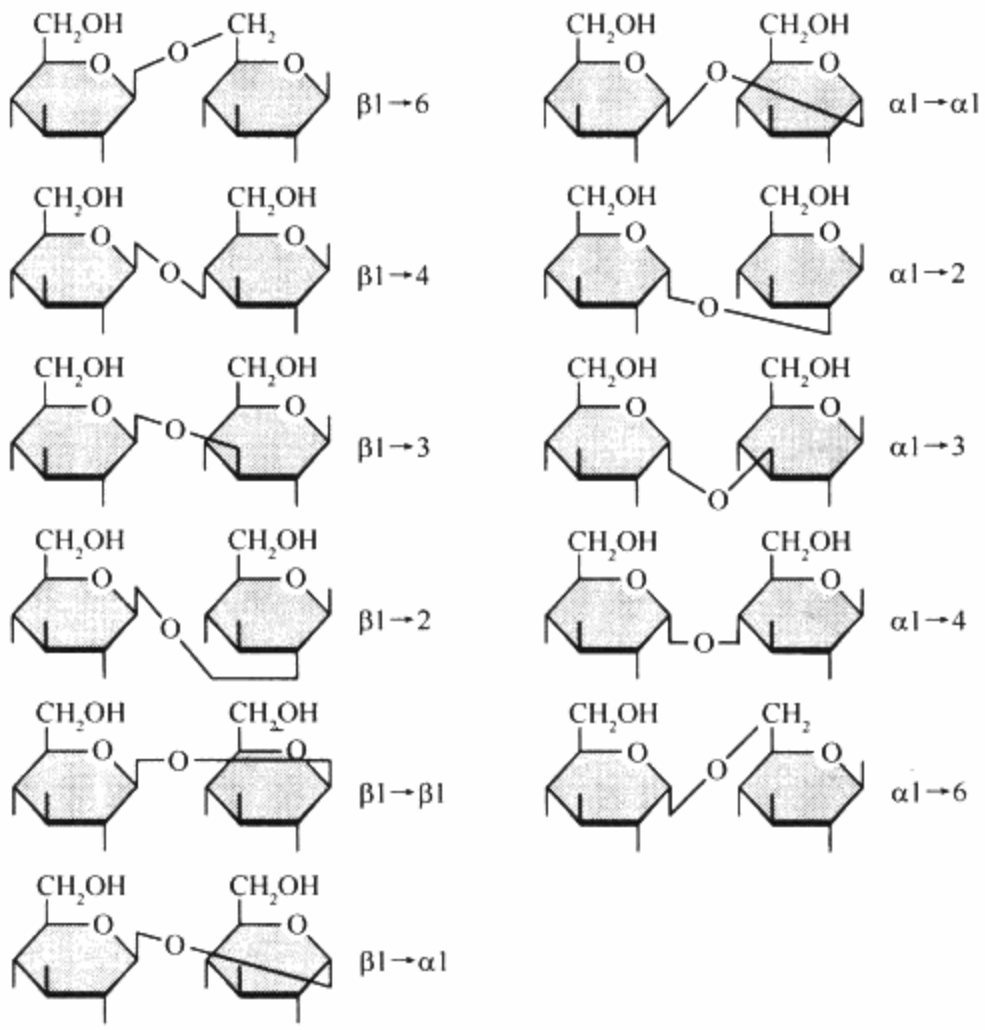


图 2-20　由两个 D-葡萄糖单位组成的 11 种二糖

尽管这些二糖仅仅在两个葡萄糖单位之间的连接键形式存在差别，它们在化学上是不同的。由于与蛋白质和脂类结合的寡糖可能拥有 6 个或者更多的不同类单糖，它们以线性或分支状排列的形式用糖苷键连接起来，如此图所示，因此细胞内可以利用的不同的寡糖种类是很多的。对于 α 和 β 连接的解释，见专题 2-4

糖不仅仅用于产生和储存能量，它们也可以用来组装机械、支撑装置。因此地球上含量最丰富的有机化学物质——植物细胞壁中的纤维素，也是葡萄糖组成的多糖。另外一种特别丰富的有机物，构成昆虫外骨骼和真菌细胞壁的壳多糖，也是一种多糖，它是一种糖衍生物，*N*-乙酰氨基葡萄糖的线性聚合物。各种各样的其他多糖是黏土、黏液、软骨的主要组分。

较小的寡糖可以同蛋白质共价结合，形成糖蛋白，也可与脂类物质结合形成糖脂，它们存在于细胞膜。如第 10 章所讲的，大部分细胞的表面都包被和连缀着糖聚合物，它们是细胞膜中的糖蛋白或者糖脂。这些糖侧链通常被其他细胞选择性地识别。人类血型不同的分子基础就在于不同人的细胞表面存在不同的糖类物质。

脂肪酸是细胞膜的组成物质

每个脂肪酸分子，例如，棕榈酸拥有两个化学上迥然不同的区域（图 2-21），一个是长长的烃链，疏水且化学反应活性较低；另一个是羧基（—COOH），它具有酸（羧酸）的特征，在溶液里电离，极其亲水，且化学反应活性较强。细胞内几乎所有的脂肪酸都通过它们的羧酸基团与其他的分子共价连接起来。

棕榈酸的烃链尾巴是饱和的：碳原子之间没有双键，含有最多数目的氢。硬脂酸，动物脂肪中另一种普通脂肪酸，也是饱和的。其他一些脂肪酸，如油酸，有不饱和的尾巴，在尾巴中存在一个或多个双键。这些双键在分子上形式一个结节，影响分子包装起来形成固体的能力。也正是因为这些双键，导致硬的（饱和的）和软的（多不饱和）人造黄油之间产生差

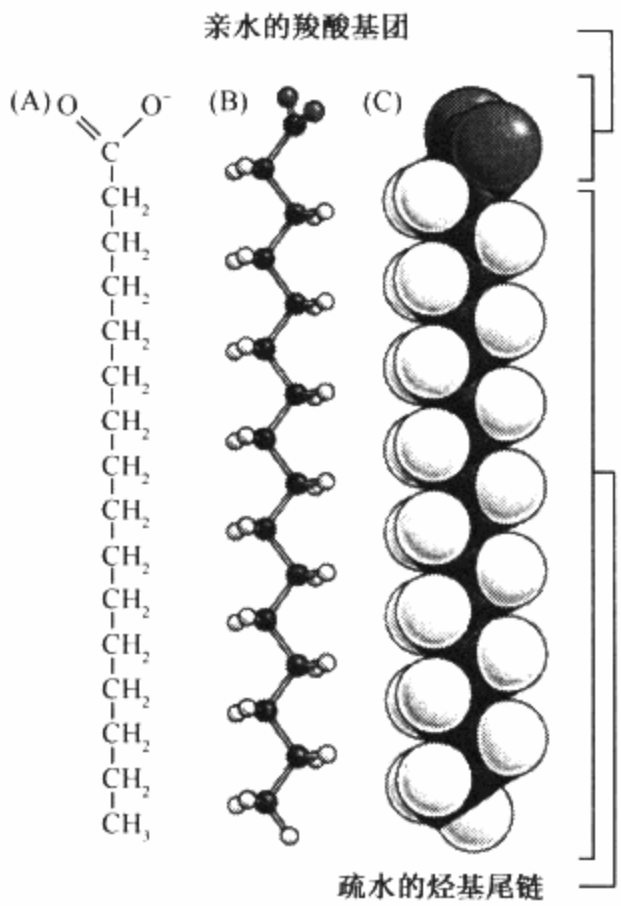


图 2-21　一种脂肪酸

脂肪酸由一个亲水的羧酸基团结合到一个疏水的烃链尾巴上产生，这里显示的是棕榈酸，不同的脂肪酸有不同的烃尾。（A）结构式。羧酸基团用电离形式表示。（B）球棒模型。（C）空间填充模型

别。细胞内发现的许多不同脂肪酸之间的差别仅仅在于烃链尾巴的长短不同和碳碳双键的数目以及位置不同（专题 2-5）。

由于脂肪酸降解能够产生的有用能量，以相同重量计算大约是葡萄糖的 6 倍，所以脂肪酸相当于细胞内浓缩的食物储存。它们以三脂酰甘油分子组成的脂肪滴的形式储存在许多细胞的细胞质内。这些三脂酰甘油通过 3 个脂肪酸链结合到一分子甘油上产生（专题 2-5），动物脂肪和植物油就属于这类分子。前者存在于肉类、黄油和乳油当中；后者如玉米油、橄榄油等。如果细胞内需要提供能量，脂肪酸链就从三脂酰甘油上释放下来并降解成许多二碳单位。这些二碳单位与葡萄糖降解衍生的二碳单位相同，它们也进入相同的产能反应途径，这些内容将在本章后面讨论。

脂肪酸和它们的衍生物如三脂酰甘油都属于脂类。脂类是一类广泛定义的生物分子的集合，这些分子具有共同的特征，它们不溶于水而溶于脂肪和有机溶剂，如苯酚。脂类典型地含有一个长长的烃尾（如在脂肪酸和异二烯中），或者含有以多种形式连接的芳香环（如在类固醇类中）。

细胞内脂肪酸最重要的功能就是形成细胞膜，这些薄片包被所有的细胞和它们内部的细胞器。细胞膜大部分由磷脂组成，它们是小分子，与三脂酰甘油类似，主要由脂肪酸和甘油结合而成。在磷脂中，甘油与两个脂肪酸链结合，而不是像三酰甘油中那样与三个脂肪酸链结合。甘油上的第三个位点结合到亲水的磷酸基团上，磷酸基团再结合到一个亲水复合物如胆碱上（专题 2-5）。因此，每一类磷脂分子具有一条由两个脂肪酸链组成的疏水尾巴和一个亲水的头部，即磷酸所在位置。这样使得磷酸具有不同于绝对疏水的三脂酰甘油的物理和化学特性。像磷脂这样，有一个亲水区域和一个疏水区域的分子，称为两亲分子。

磷脂形成细胞膜的功能来源于它们的两亲特性。磷脂在水面展开会形成单层的磷脂分子，疏水尾巴朝向空气，而亲水头部接触水面。在水中，两层这样的分子层很容易尾对尾结合形成一个磷脂三明治，或者脂双层结构。这个双层结构是所有细胞膜的结构基础（图 2-22）。

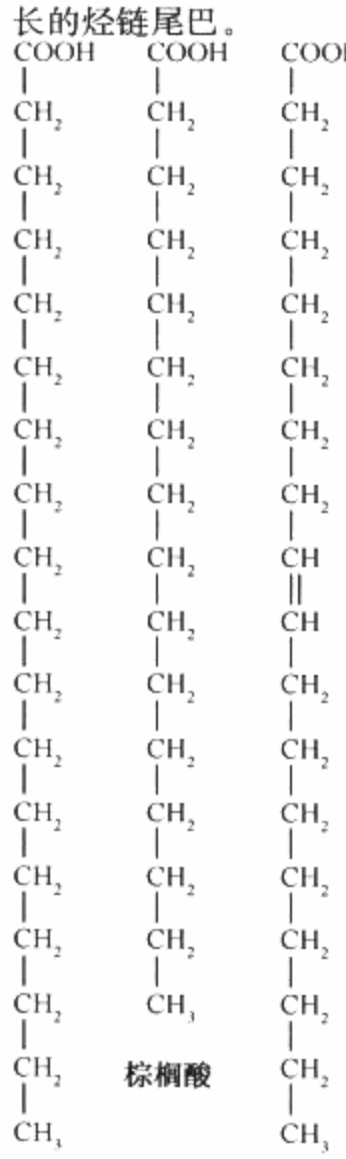
氨基酸是蛋白质的亚单位

氨基酸是一类具有一个确定特性的多种分子集合：它们都有一个羧酸基团和一个氨基，这两个基团都结合到单一的碳原子上，称为 α-碳原子（图 2-23）。氨基酸的化学多样性来源于侧链，它也结合到 α-碳原子上。氨基酸之所以对于细胞很重要，是因为它们能组成蛋白质。蛋白质就是氨基酸的多聚体，这些氨基酸头尾相连，形成一条长链，然后折叠成每种蛋白特有的三维结构。蛋白质链中相邻两个氨基酸之间的共价键称为肽键；氨基酸形成的长链称为多肽链（图 2-24）。无论多肽链由哪些特定的氨基酸组成，它的一端都有一个氨基（NH_2）基团（它的 N 端），而另一端都有一个羧基（COOH）基团（它的 C 端）。这使得它具有确定的方向——结构上的极性（与电极相反）。

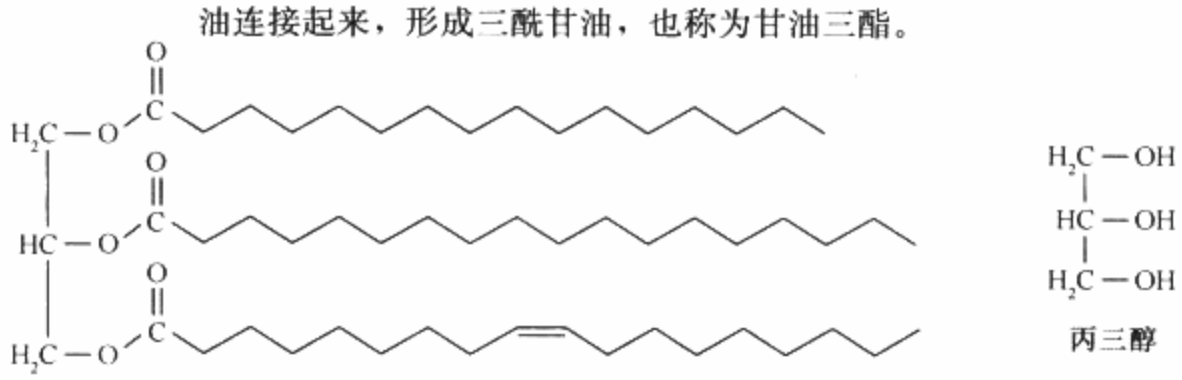
专题 2-5　脂肪酸和其他的脂类

常见的脂肪酸

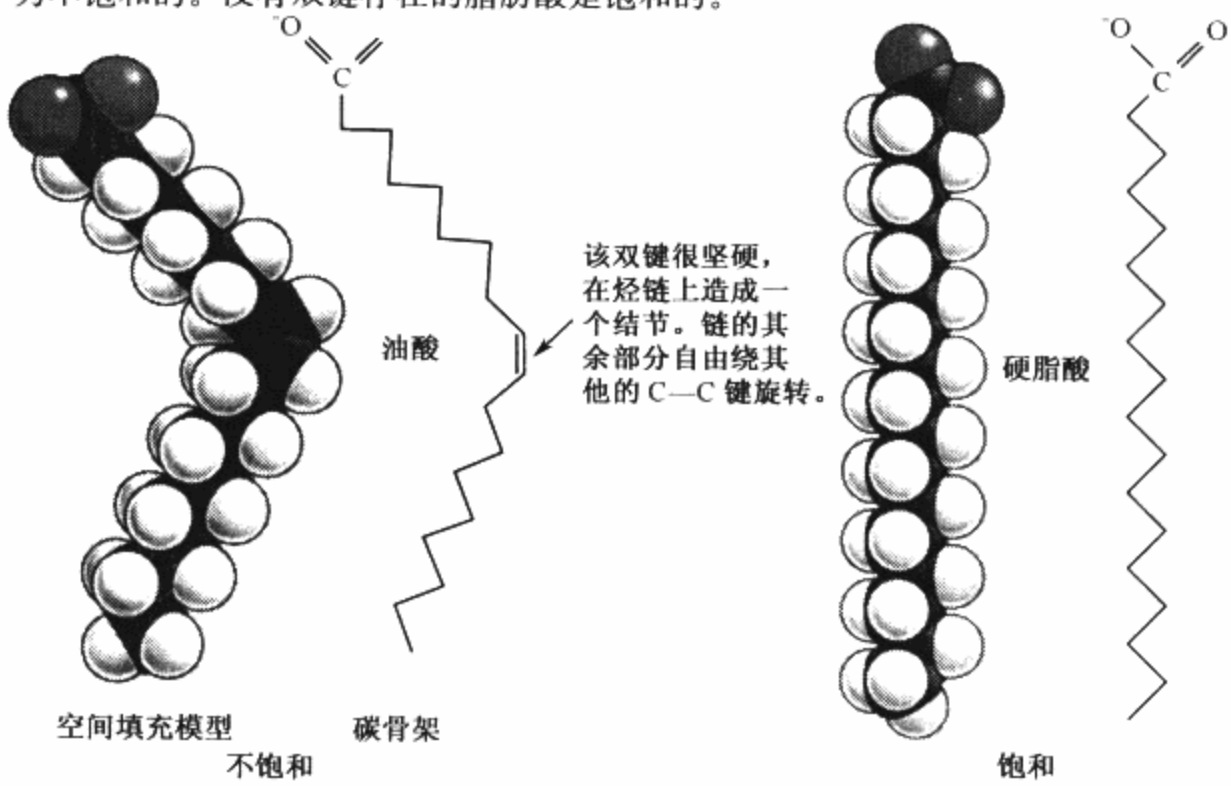
这些都是羧酸，具有长长的烃链尾巴。



三酰甘油　脂肪酸被当作能量（脂肪和油）储存起来，它通过一个酯键与甘油连接起来，形成三酰甘油，也称为甘油三酯。

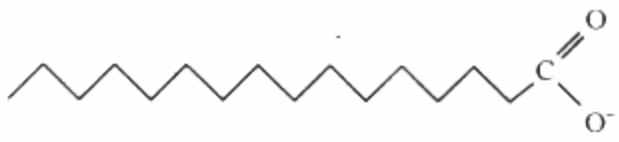


存在数百种不同的脂肪酸。其中一些脂肪酸的烃尾中有一个或多个双键，称其为不饱和的。没有双键存在的脂肪酸是饱和的。

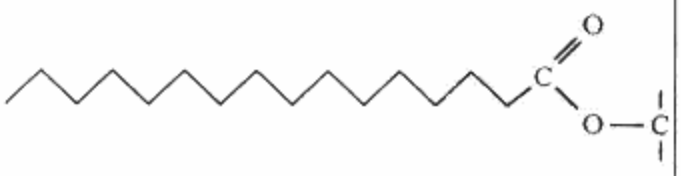


羧基

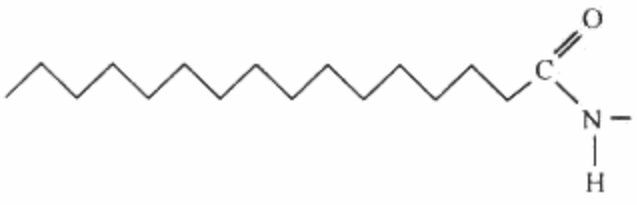
如果以游离的形式存在，脂肪酸的羧基就会电离。



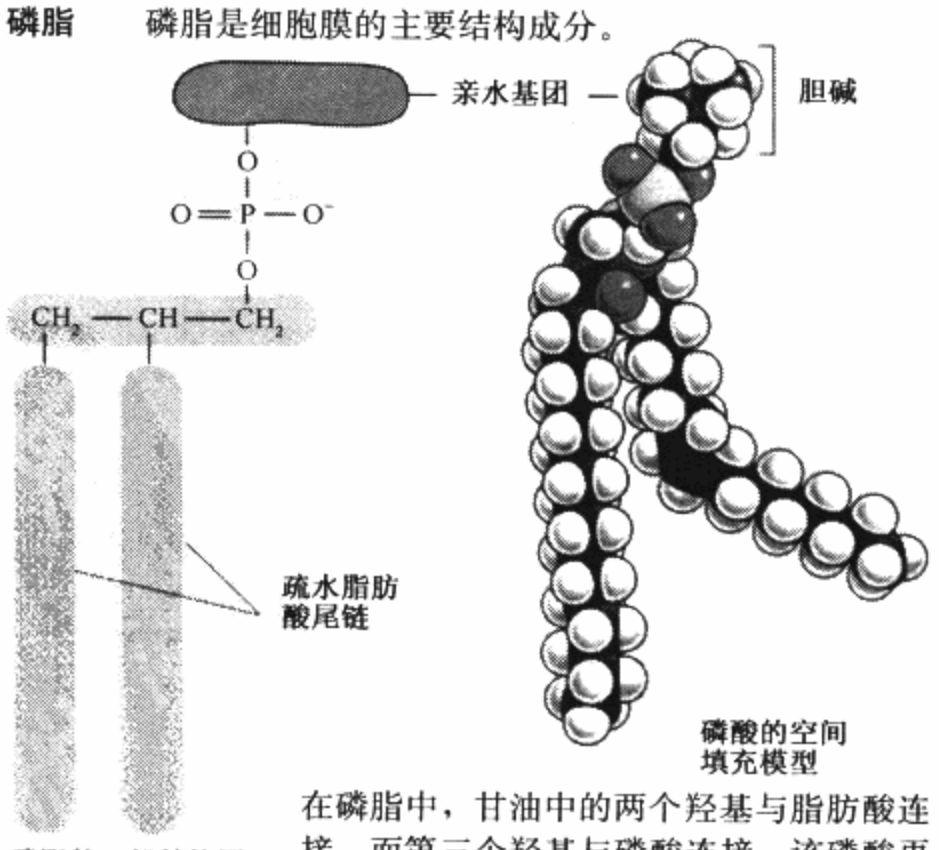
但多数情况下羧基与其他基团结合，形成酯



或酰胺。

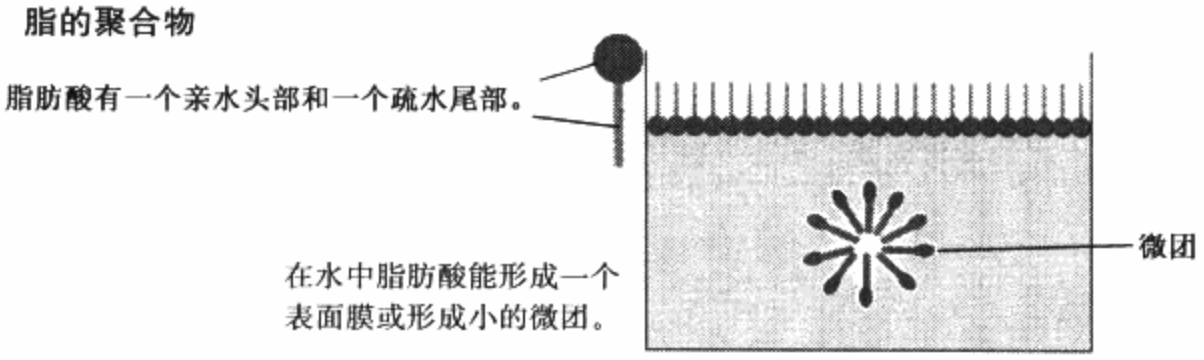


磷脂　磷脂是细胞膜的主要结构成分。



在磷脂中，甘油中的两个羟基与脂肪酸连接，而第三个羟基与磷酸连接。该磷酸再与一个较小的极性基团（乙醇）连接。

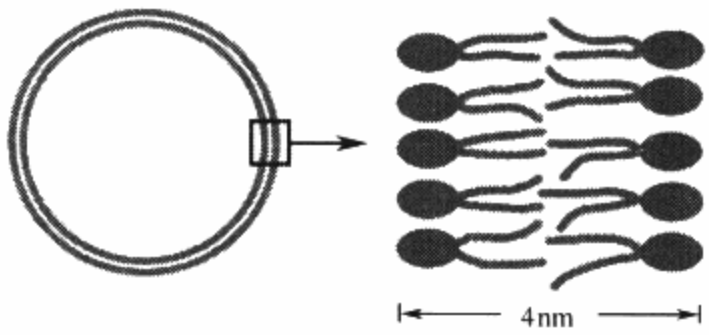
脂的聚合物



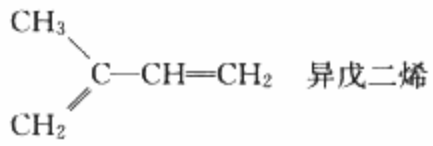
它们的衍生物可以通过疏水作用力形成更大的聚合物：

三酰甘油可以在细胞质内形成巨大的球状脂肪滴。

磷脂和甘油酯形成自我封闭的脂双层作为细胞膜的基础。

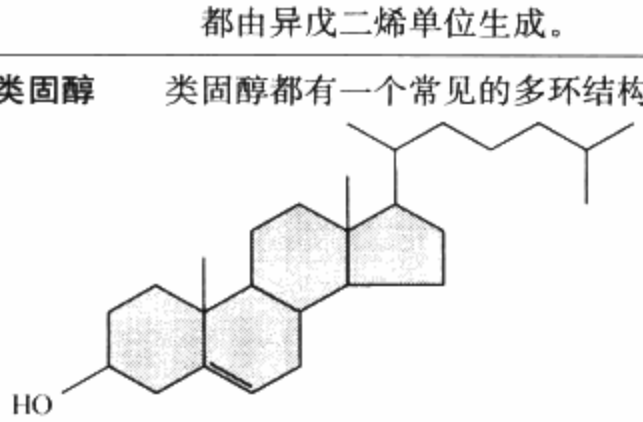


其他的脂质 脂质是指细胞内不溶于水而溶于有机溶剂的分子。其他两类常见的脂是类固醇和多聚类异戊二烯，二者都由异戊二烯单位生成。

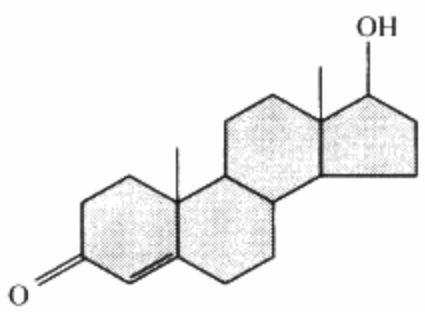


异戊二烯

类固醇 类固醇都有一个常见的多环结构。



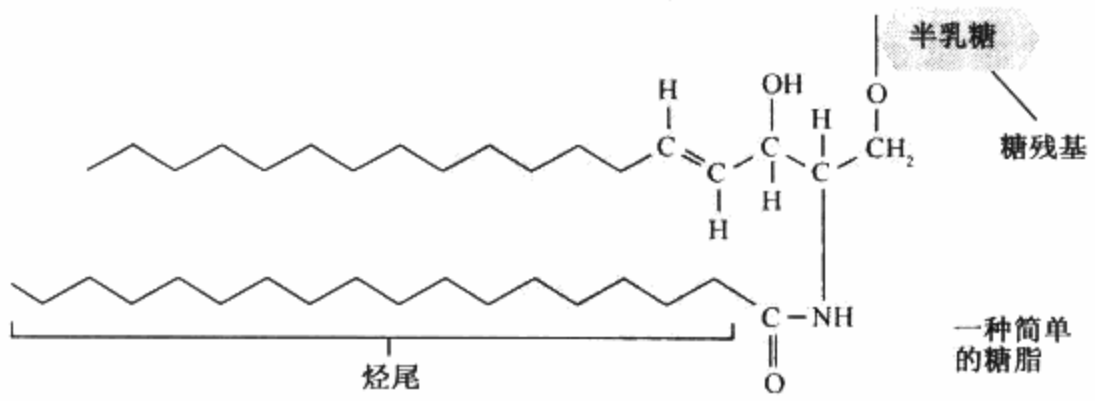
胆固醇——存在于许多细胞膜中



睾酮——男性胆固醇激素

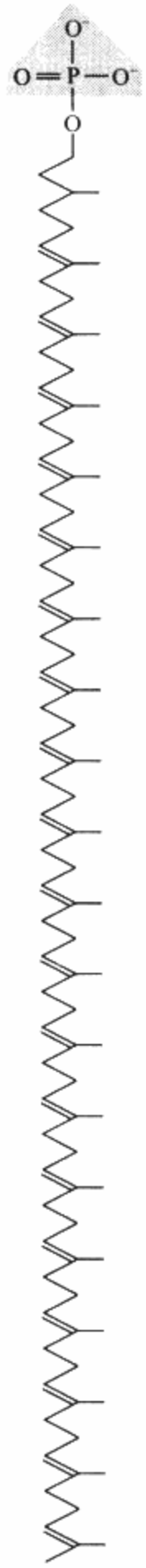
糖脂

与磷脂一样，这些复合物由疏水区和极性区组成，其中疏水区含有 2 个长长的烃尾，而极性区域则含有一个或多个糖残基，不含磷酸。



多聚类异戊二烯

异戊二烯的长链复合物



多萜醇磷酸——用于在有细胞膜参与多糖蛋白和某些多糖的合成中携带活化的糖类物质。

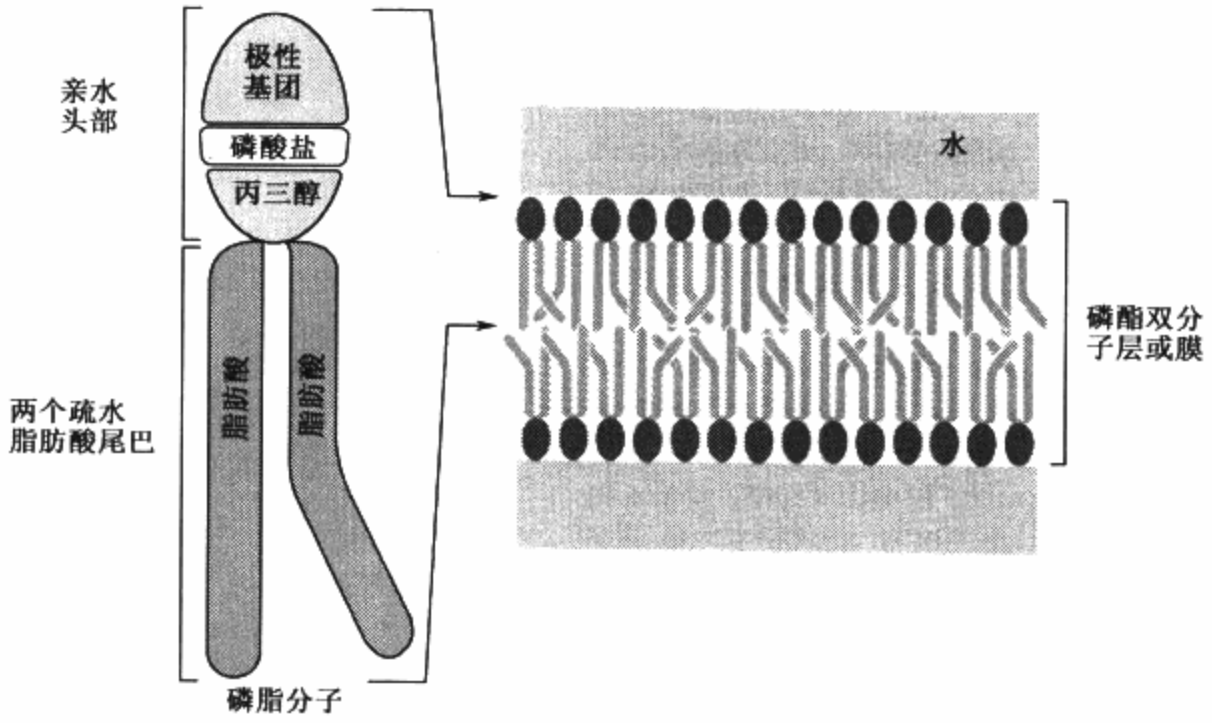


图 2-22　磷脂结构以及磷脂在细胞膜内的方向性

在水环境中，磷脂的疏水尾巴聚集在一起排斥水。这里它们形成了一个双层结构，每个磷脂的亲水头部朝向水。脂双层是细胞膜的基础，这些内容将在第 10 章详细讨论

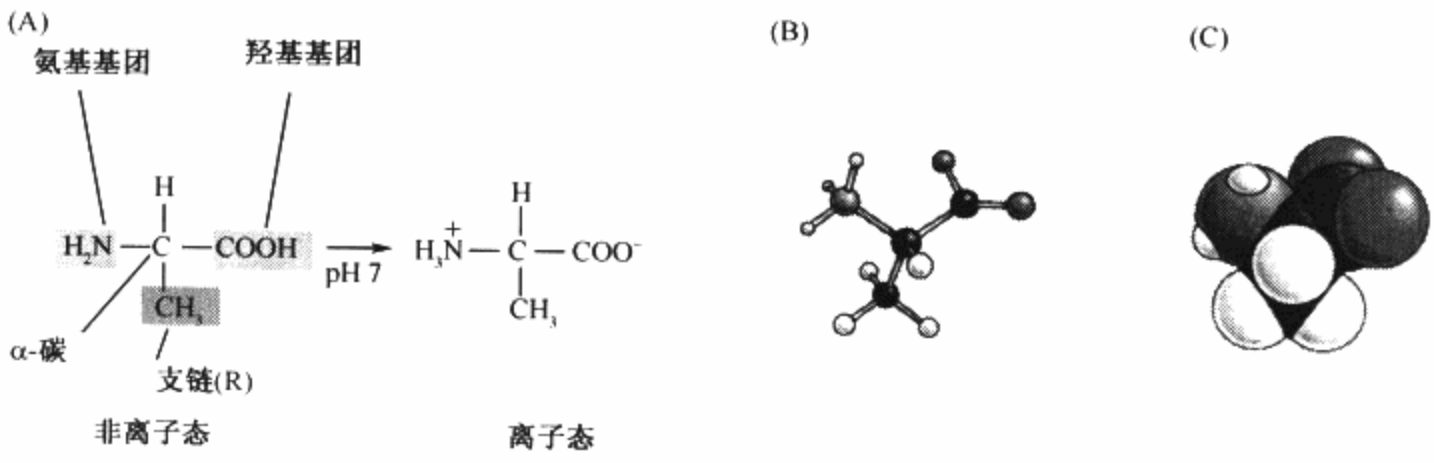


图 2-23　丙氨酸

（A）在细胞内，pH 值接近 7，游离的氨基酸以电离形式存在，但是它一旦渗入到多肽链中，氨基和羧基所带的电荷就会消失。（B）丙氨酸的球棒模型。（C）丙氨酸的空间填充模型（H，白色；C，墨色；O，红色；N，蓝色）

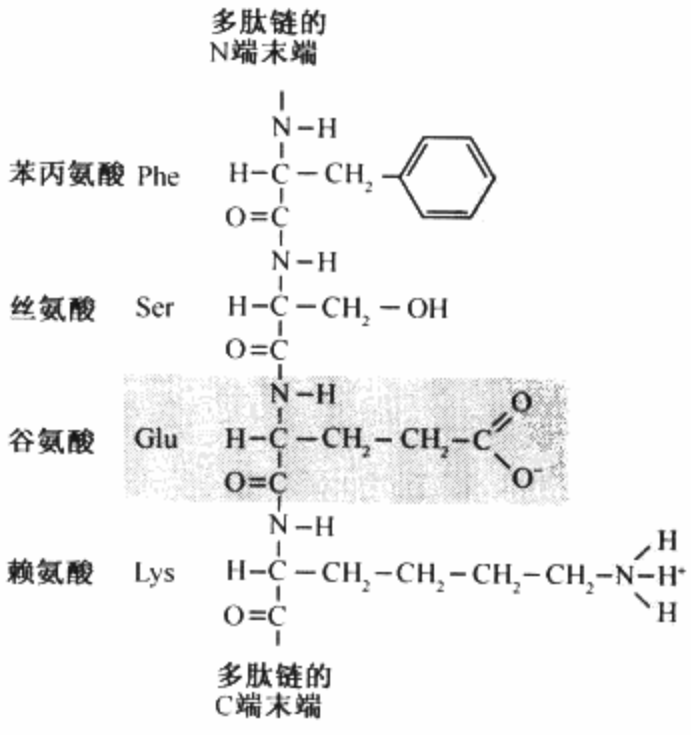


图 2-24　蛋白质分子的一小部分

所示的 4 个氨基酸通过 3 个肽键连接起来，其中一个肽链用黄色突出表示。其中一个氨基酸用灰色阴影表示。氨基酸侧链用红色表示，多肽链的两个末端在化学上是不同的。一个末端，N 端，是氨基；另一个末端，C 端，是羧基。序列通常从 N 端开始阅读，因此这个序列是 Phe-Ser-Glu-Lys

蛋白质中通常存在 20 种氨基酸，每一种都有不同的侧链结合到 α-碳原子上（专题 3-1）。这 20 种氨基酸在所有的蛋白质中不断出现，而不论这些蛋白质是来自于细菌、植物还是动物。生命进化过程中，这 20 种氨基酸是如何被选定作为蛋白质组分的还不清楚；其他氨基酸为什么不能同样好地发挥作用的化学原因也不清楚。但是一旦确立了这种选择结果，就不能再更改了，因为生命的太多方面依赖于它。

与糖一样，所有的氨基酸除甘氨酸之外，都存在旋光异构体，*D* 型或 *L* 型（专题 3-1）。但是蛋白质中仅仅发现 *L* 型的存在，尽管 *D* 型氨基酸是细菌细胞壁和一些抗生素的一部分。这种仅使用 *L* 型而不使用 *D* 型构造蛋白质的原因是另一个尚未弄清楚的进化之谜。

20 种标准氨基酸提供的化学多样性对于蛋白质功能是至关重要的。这 20 种氨基酸有 5 种氨基酸的侧链在溶液中能电离，因此能带上电荷（图 2-25）。其他的是不带电荷的氨基酸。

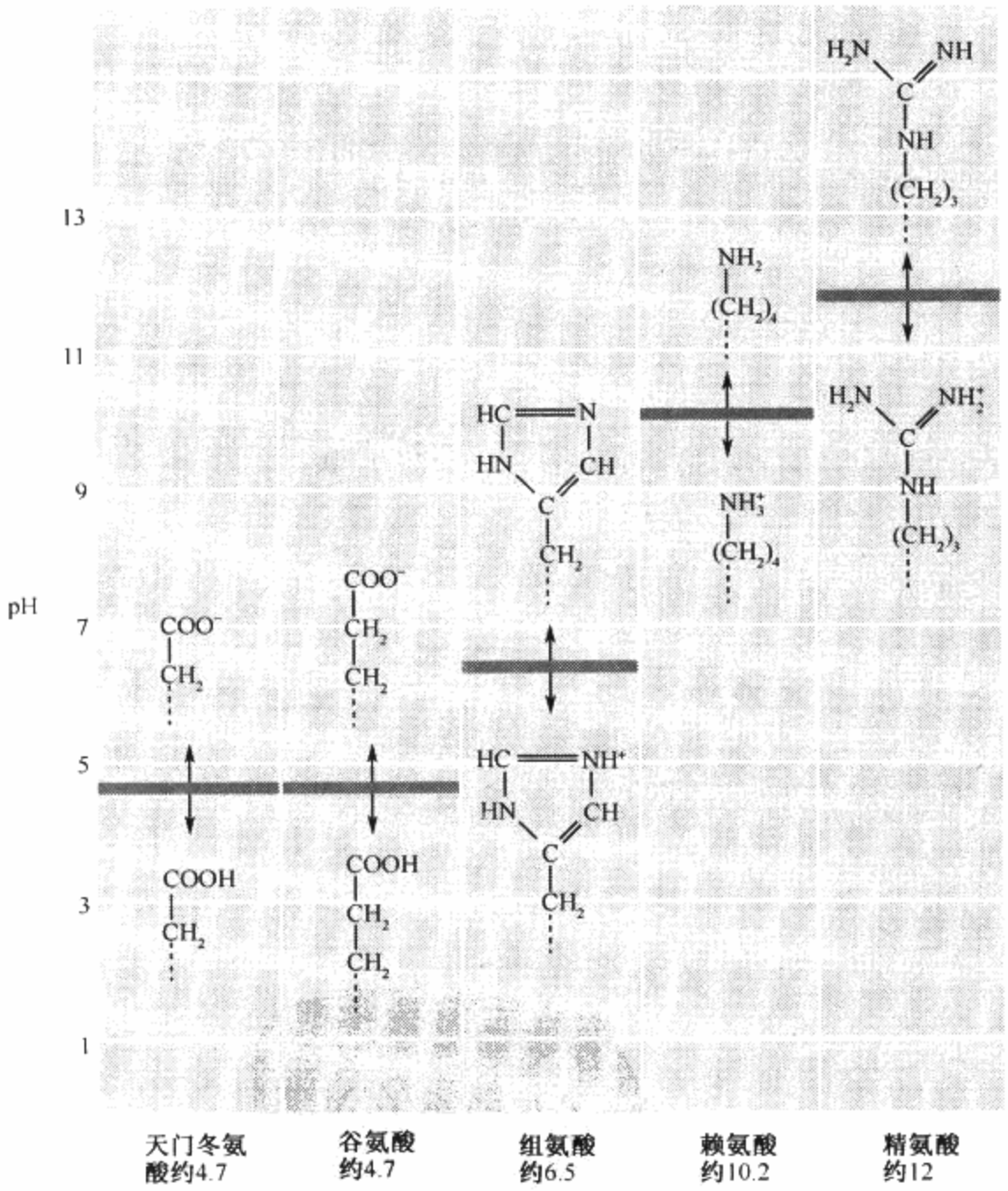


图 2-25　氨基酸侧链所带电荷依赖于 pH 值

图中显示了 5 种不同的能带电荷的侧链。在水溶液中，羧酸很容易失去 H^+，形成带负电荷的离子，可用词缀“-ate”表示，如天冬氨酸和谷氨酸中那样。对于氨基，也存在类似的情况，在水溶液中氨基酸可以接受一个 H^+ 形成带正电荷的离子（没有特殊的名称）。这些反应是迅速可逆的，带电荷与不带电荷这两种形式的数目依赖于溶液的 pH 值。高 pH 值时，羧酸倾向于带电荷，而氨基倾向于不带电荷，低 pH 值时则相反，羧酸不带电荷，而氨基带电荷。当半数的羧酸或氨基残基带电荷时的 pH 值称为这个氨基酸侧链的 pK。在细胞中，pH 接近 7，几乎所有的羧酸和氨基都以带全部电荷的形式存在

它们中一些是极性和亲水的，另一些是非极性和疏水的。正如我们第 3 章将讨论的，氨基酸侧链的多种特性是蛋白质具有多样且复杂的功能的基础。

核苷酸是 DNA 和 RNA 的亚单位

核苷酸是由一个含氮环复合物与一个 5 碳糖结合形成的一类分子。这个糖可以是核糖，也可以是脱氧核糖，它携带一个或更多的磷酸基团（专题 2-6）。含有核糖的核苷酸也称为核糖核苷酸，而含有脱氧核糖的也称为脱氧核糖核苷酸。由于历史原因，含氮环通常称为碱基：酸性条件下，它们每一个可以结合一个 H^+（质子），因此能增加水溶液中 OH^- 浓度。不同的碱基之间具有很强的家族相似性。胞嘧啶（C）、胸腺嘧啶（T）和尿嘧啶（U）之所以称为嘧啶，是因为它们都是含有 6 个原子的嘧啶环的衍生物；鸟嘌呤（G）和腺嘌呤（A）是嘌呤复合物，它们含有另外一个五原子环，与六原子环融合。每个核苷酸都是根据它所含的碱基命名的（专题 2-6）。

核苷酸可以作为化学能的短期载体，其中最重要的是**腺嘌呤核苷三磷酸**或 **ATP**（图 2-26），它用于成百上千的不同细胞反应之间转移能量，ATP 是在食物的氧化降解释放的能量，推动进行的反应中生成的。它的 3 个磷酸基团通过 2 个磷酸酐键依次相连，这些磷酸酐键的断裂会释放出大量有用的能量。尤其是末端的磷酸基团，频繁地被水解断裂，将一个磷酸基团转移到其他分子上或者释放能量推动需能生物合成反应（图 2-27）。其他的核苷酸衍生物可以作为其他化学基团的转移载体，这些将在后面讨论。

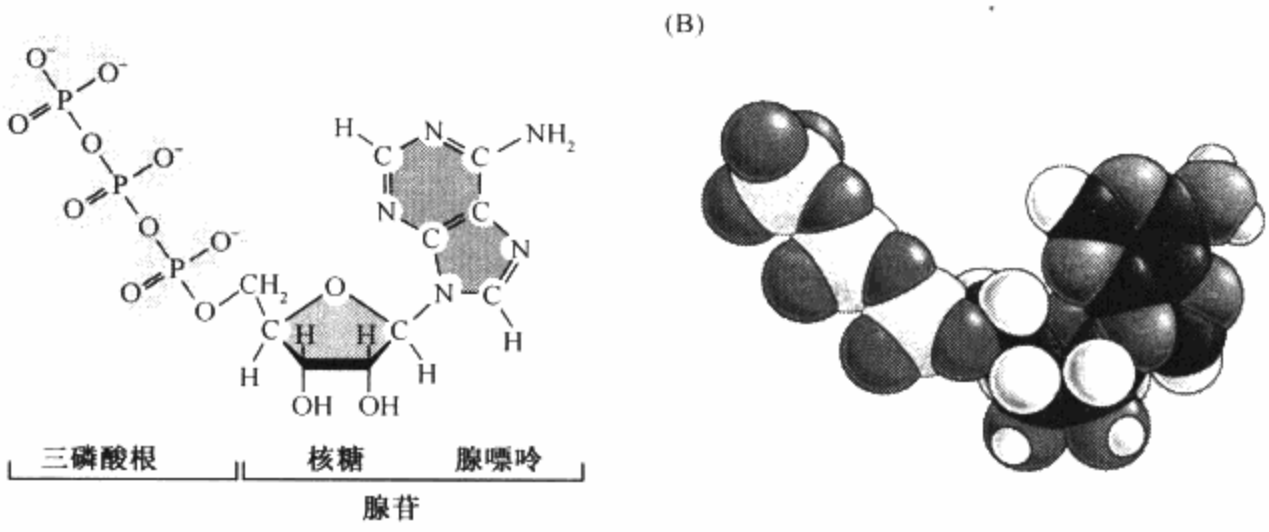


图 2-26　腺嘌呤核苷三磷酸（ATP）的化学结构

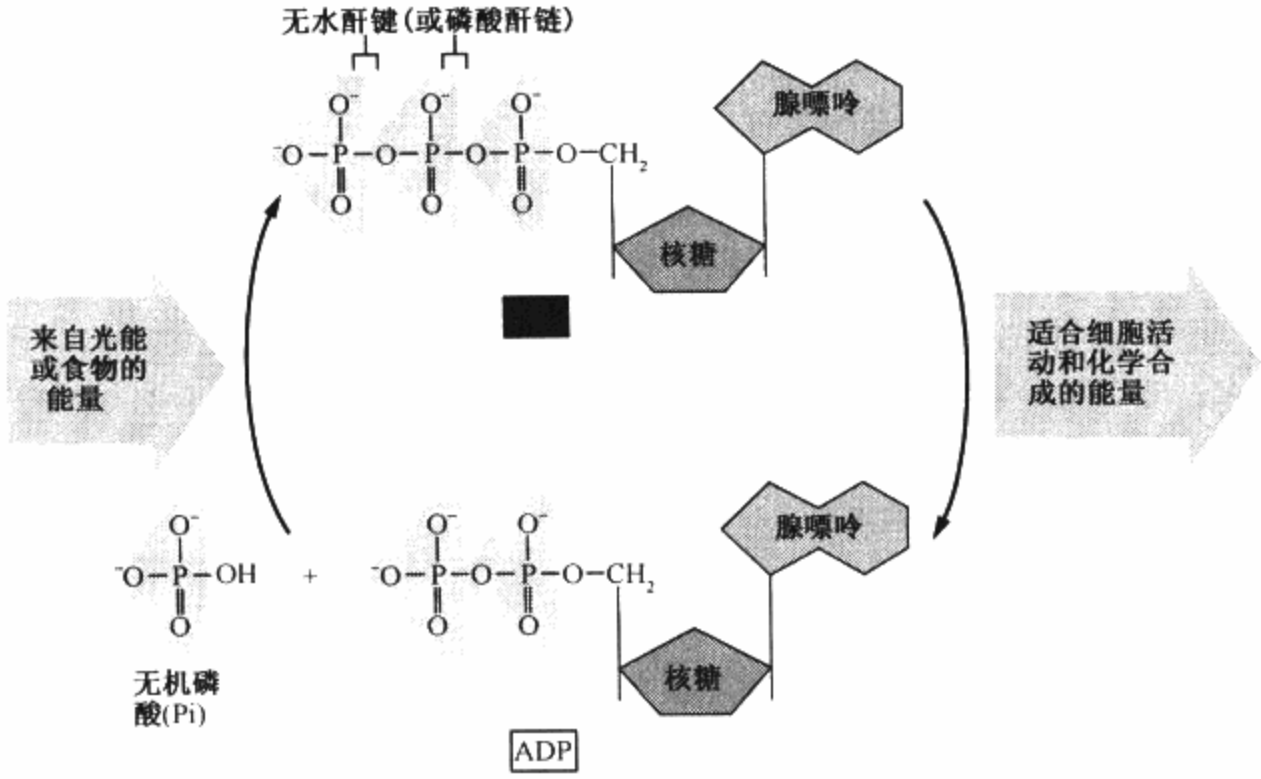
（A）结构式；（B）空间填充模型；（B）中代表原子的颜色分别是：C，黑色；N，蓝色；H，白色；O，红色，以及 P 黄色

然而，核苷酸在细胞内最基本的作用是储存以及提取生物信息。核苷酸作为构建单元，用于形成核酸——它是长链的聚合物，其中核苷酸亚单位之间通过磷酸二酯键共价连接起来，该键是在上一个核苷酸中结合到核糖上的磷酸基团和下一个核苷酸中核糖上的羟基之间生成的。核酸链是由富含能量的核苷三磷酸经过缩合反应合成的，这个缩合反应在磷酸二酯键形成过程中，释放出无机磷酸。

存在两种主要类型的核酸，它们的差别在于糖-磷酸骨架上，存在不同类型的糖。那些以核糖为基础的核酸称为**核糖核酸**（ribonucleic acid）或 **RNA**，它们含有 A、G、C 和 U 碱基。而那些以脱氧核糖（分子中核糖碳原子环上 2′位置的羟基被氢取代，见专题 2-6）为基础的核酸称为**脱氧核糖核酸**（deoxyribonucleic acid）或 **DNA**，它们含有 A、G、C 和 T 碱基（T 在

图 2-27　ATP 分子是细胞内的能量载体

由 ADP 和无机磷酸到 ATP 的需能合成过程与产生能量的食物的氧化（动物细胞、真菌和一些细菌中）或者光能捕获（植物细胞和一些细菌中）过程偶联起来。反过来，ATP 水解产生 ADP 和无机磷酸的过程则提供能量推动许多细胞反应



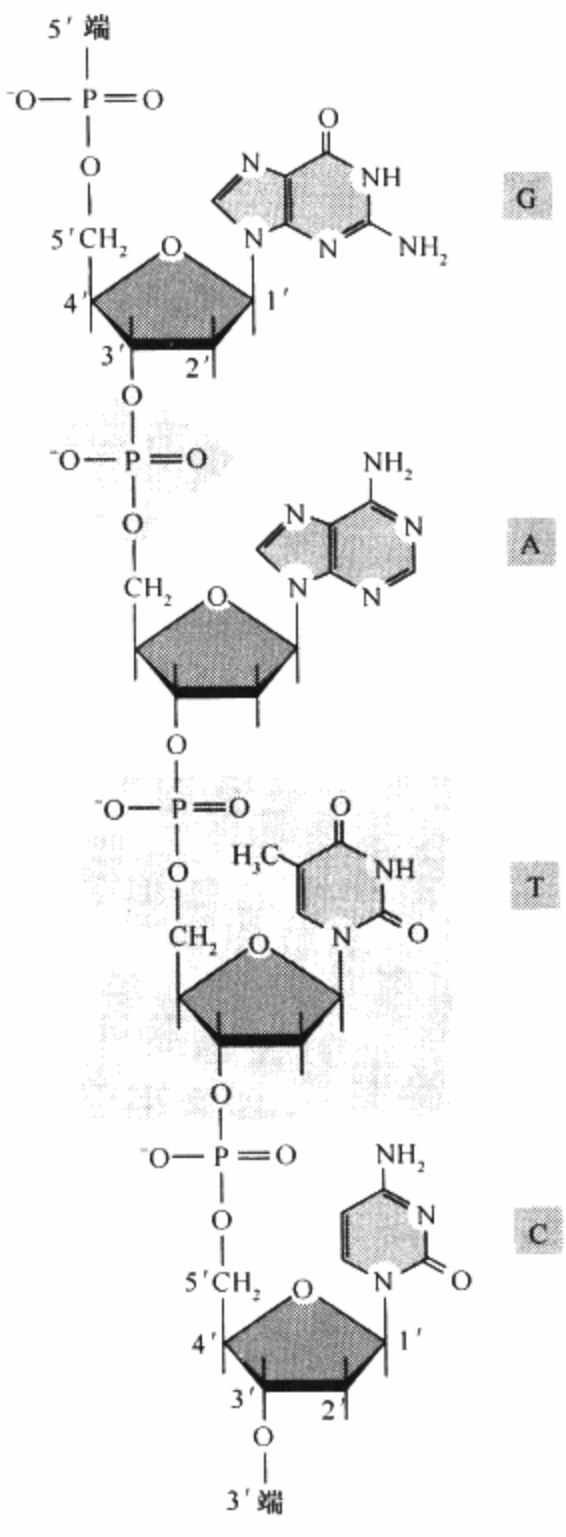


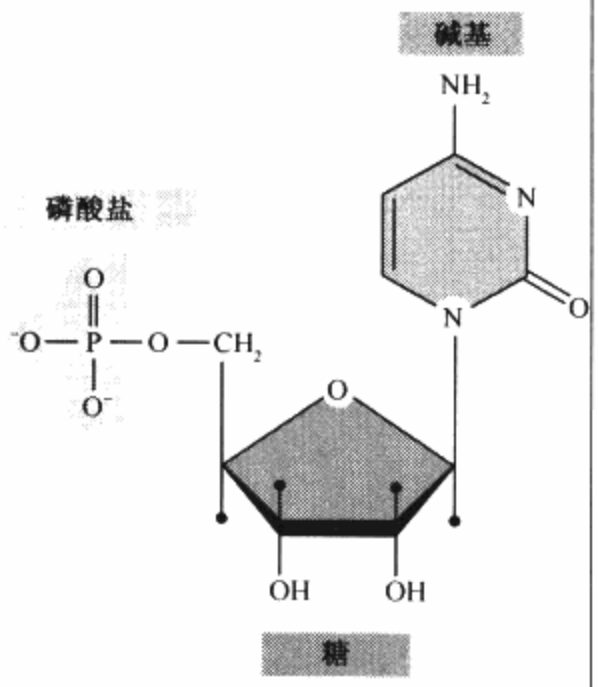
图 2-28　脱氧核糖核酸（DNA）分子一条链的一小部分

图中显示了 4 个核苷酸，用黄色突出显示了其中一个相邻核苷酸之间的磷酸二酯键，用灰色阴影显示了其中一个核苷酸。核糖上特定位置的碳原子称为 5′和 3′碳原子，之间的磷酸二酯键将核苷醇连接起来。因为如此，多聚核苷酸链的一端，5′端，具有一个自由磷酸基团，而另一端，3′端，具有一个羟基。多聚核苷酸链中核苷酸的线性序列通常用一个字母的代码进行缩写，而且序列总是从 5′端开始阅读。本例所示的序列是 G-A-T-C

专题 2-6　核苷酸的概括

核苷酸

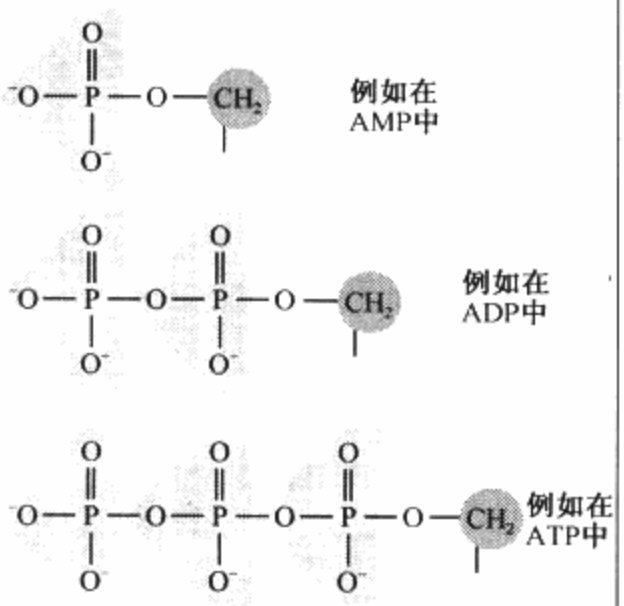
每个核苷酸分子含有一个含氮碱基、一个五碳糖，以及一个或多个磷酸基团。



核苷酸是核酸的基本单位。

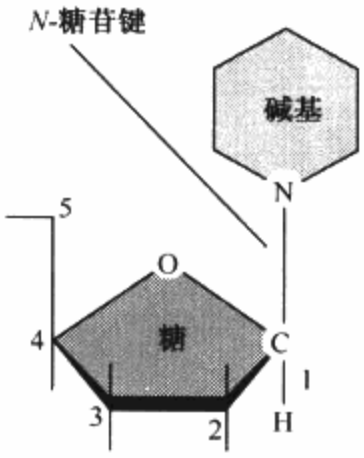
磷酸

磷酸通常与核糖或脱氧核糖（5′）的C5位羟基相连。单磷酸、二磷酸和三磷酸都比较常见。



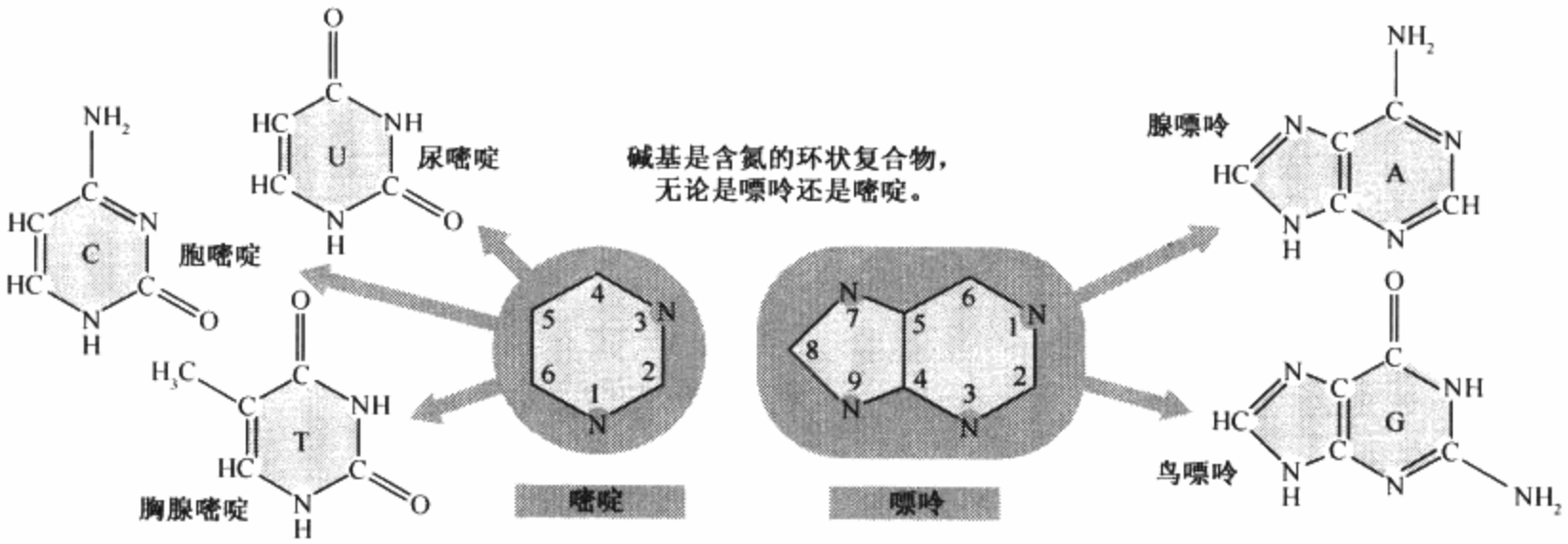
磷酸使核苷酸带负电荷。

基本的糖苷键



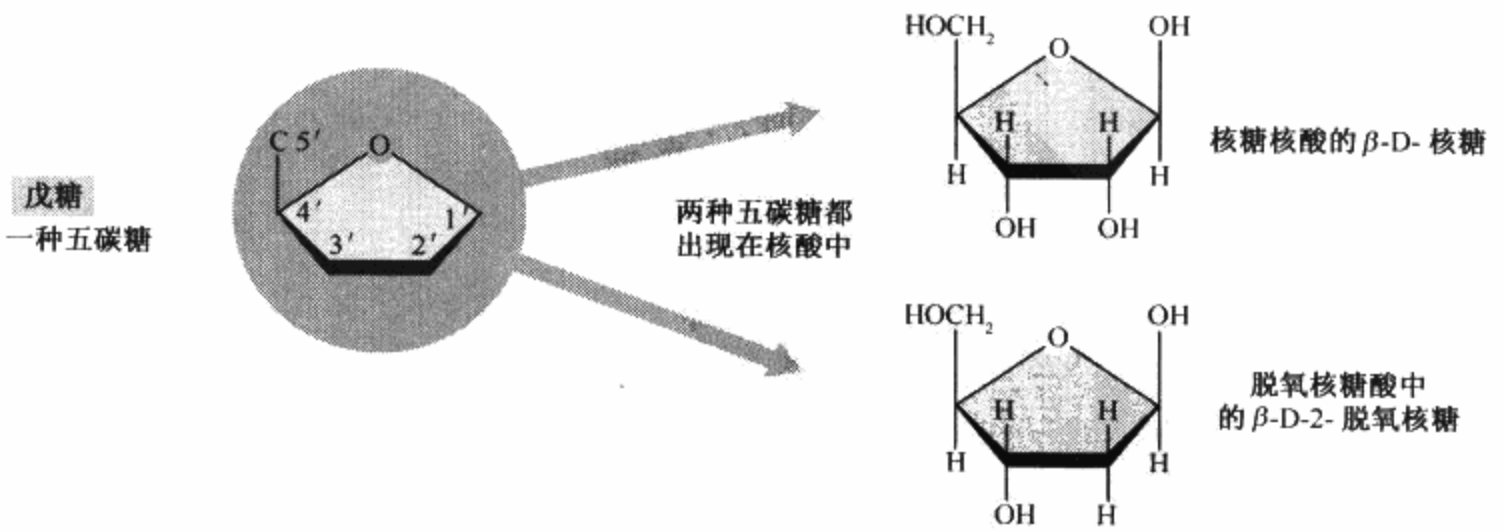
碱基与同一碳原子（C1）以糖-糖键相连。

碱基



碱基是含氮的环状复合物，无论是嘌呤还是嘧啶。

糖



核苷酸的糖的每一个标记的碳原子都加撇表示；因此，其中一个读作“5 撇碳”等。

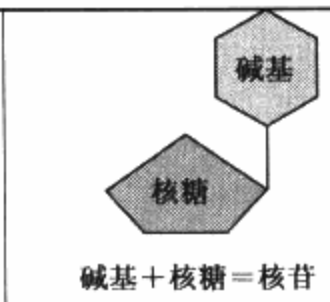
术语

全名称可能会混淆不清，缩写就很清楚了

碱基	核苷	缩写
腺嘌呤	腺苷	A
鸟嘌呤	鸟嘌呤核苷酸	G
胞嘧啶	胞嘧啶核苷酸	C
尿嘧啶	尿嘧啶核苷酸	U
胸腺嘧啶	胸腺嘧啶核苷酸	T

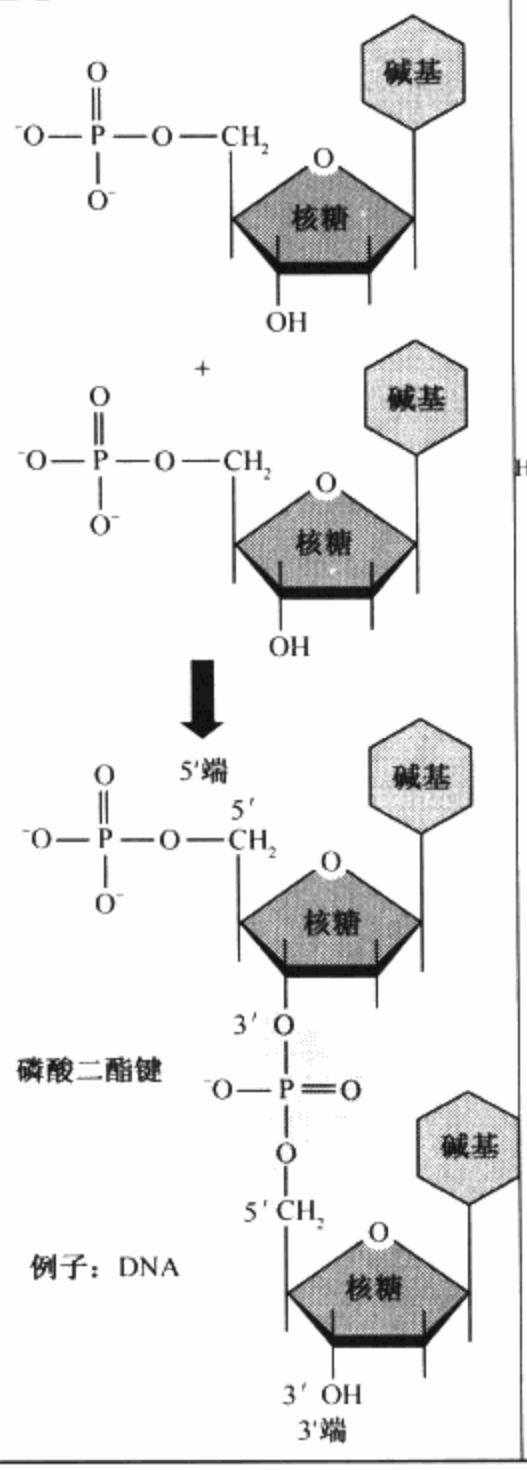
核苷酸可用三个字母简写，下面是一些例子：

AMP＝腺嘧呤核苷单磷酸
dAMP＝脱氧腺嘧呤核苷单磷酸
UDP＝尿嘧啶核苷二磷酸
ATP＝腺嘌呤核苷三磷酸



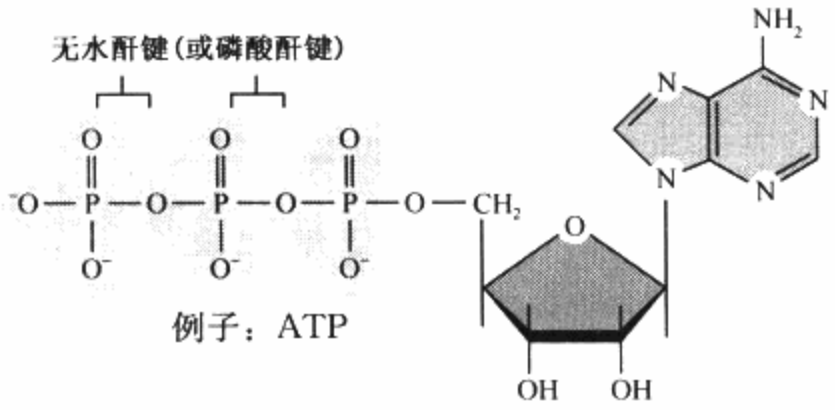
核酸

核苷酸通过 5′和 3′碳原子之间的磷酸二酯键连接起来，形成核酸。核酸链中核苷酸的线性序列通常用一个字母的符号简写，A-G-C-T-T-A-C-A，并将链的 5′端写在左边。

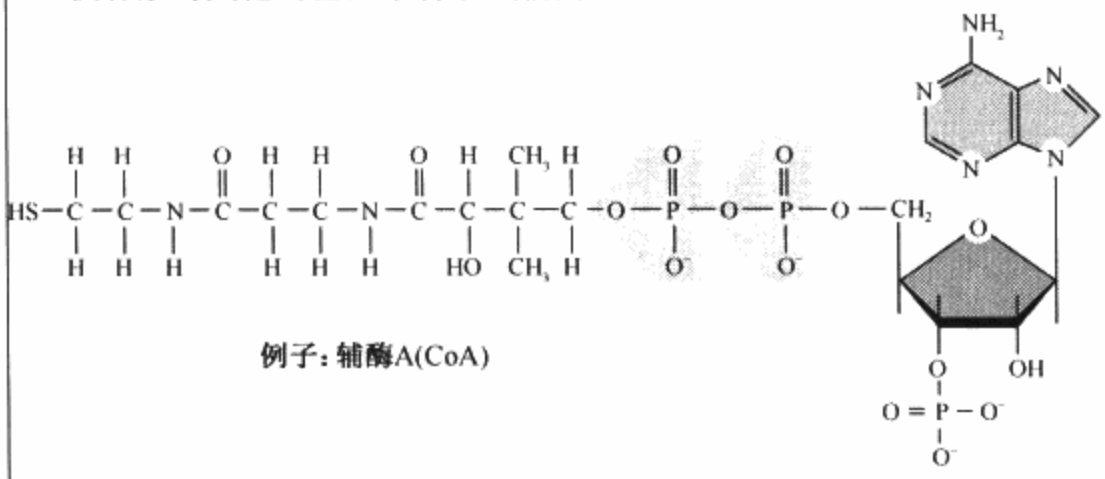


核苷酸具有许多其他的功能

1. 核苷酸可以通过极易水解的磷酸酐键携带化学能量。

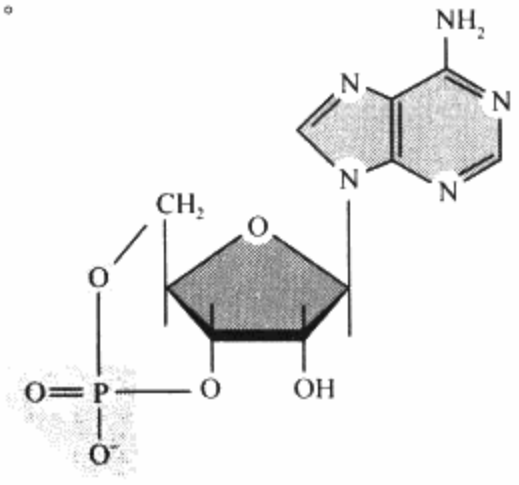


2. 核苷酸与其他的基团结合形成辅酶。



3. 核苷酸用作细胞内的特殊信号分子。

例子：环腺苷磷酸（cAMP）



化学上类似 RNA 中的 U，仅仅在嘧啶环上增加一个甲基）（图 2-28）。RNA 在细胞中通常以一条多聚核苷酸链的形式存在，而 DNA 实际上总是以双链分子存在，两条走向相反的多聚核酸链组成一个 DNA 双螺旋，它们通过在两条链的碱基之间形成氢键彼此缠绕在一起。

DNA 或 RNA 中核苷酸的线性序列编码了细胞的遗传信息。不同核酸分子中碱基通过氢键彼此识别和配对（称为碱基对）的能力，即 G 和 C 配对，A 和 T 或 U 配对，是所有遗传和进化的基础，这些将在第 4 章讨论。

拥有显著特征的大分子在细胞化学中占据主要地位

以重量衡量，大分子是目前为止活细胞内最为丰富的含碳分子（图 2-29 和表 2-4）。它们是组装细胞的主要构成元件，也是赋予了生命体不同特征的组分。细胞内的大分子是有机小分子（称为单体或亚单位）仅仅通过共价连接成一条长链形成的多聚体（图 2-30）。然而这些大分子拥有许多显著的特征，这些特征是不可能从它们的简单组分中预测到的。

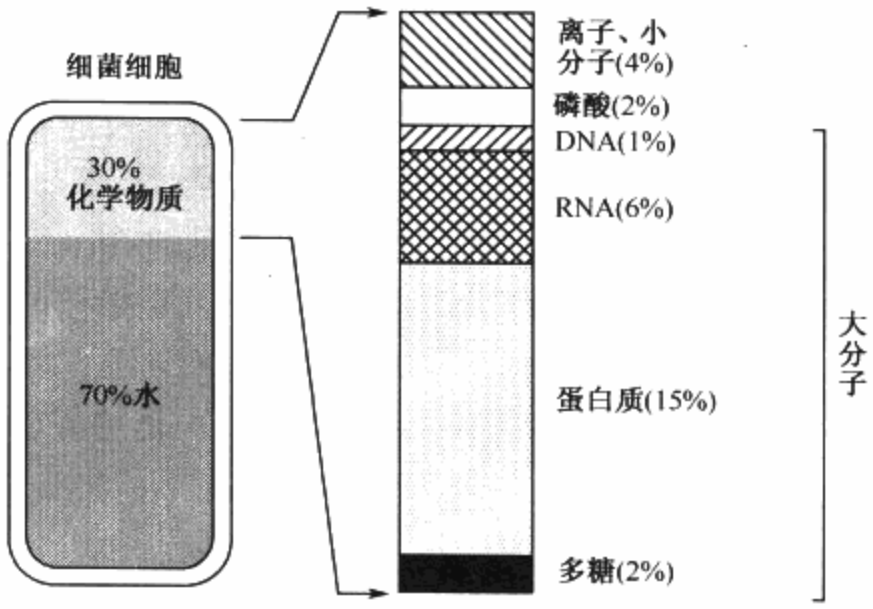


图 2-29　细胞内大分子含量丰富

图中显示细菌细胞大致组成（以重量计算）。动物细胞的组成与其类似（表 2-4）

表 2-4　一个典型的细菌和一个典型的哺乳动物细胞近似的化学组成

成分	占细胞比重	
	大肠杆菌	哺乳动物细胞
水	70	70
无机离子(Na^+，K^+，Mg^{2+}，Ca^{2+}，Cl^-等)	1	1
各种小分子代谢产物	3	3
蛋白质	15	18
RNA	6	1.1
DNA	1	0.25
磷酯	2	3
其他脂类	—	2
多糖	2	2
细胞总体积	$2\times10^{-12}cm^3$	$4\times10^{-9}cm^3$
相对细胞体积	1	2000

蛋白质多糖、DNA 和 RNA 是生物大分子。脂类虽然具有大分子的某些特征。但是并不被归为生物大分子；例如：脂类大部分被合成为小分子的线性多聚体（乙酰辅酶 A 上的乙酰基团），同时它们自我聚集成大的结构（如细胞膜）。注意水和蛋白质构成了细菌和动物细胞的大部分质量。

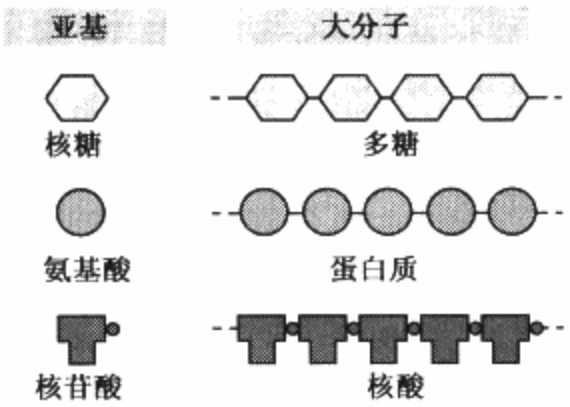


图 2-30　三类大分子

每种大分子都是由小分子（称为单体或亚单位）通过共价键连接起来形成的聚合物

特别是蛋白质含量丰富，种类多样。它们在细胞内发挥无以计数的各种作用。许多蛋白质作为酶，催化细胞所需的大量共价键生成和断裂的反应。所有细胞借以从食物分子中获取能量的反应都由作为酶的蛋白质进行催化，如一种核糖二磷酸羧化酶。在光合作用有机体内将 CO_2 转变为糖，生成地球上生命所需的大部分有机物质。其他一些蛋白质用于构造结构组分，例如，微管蛋白，它是一种可以自我组装成细胞长长的微管的蛋白质，或者组蛋白，它们是一类将 DNA 压缩成染色体的蛋白质。但是也有一些蛋白质充当分子泵来产生力量和位移，如肌肉中的肌球蛋白。蛋白质也可以拥有大量的其他功能。在本书后面我们将讨论这些功能产生的分子基础。这里我们仅仅提及使得这些功能成为可能的大分子化学中的一些共同原则。

尽管从细节上讲，对于将单体加到每个聚合物上生成蛋白质、核酸和多糖的反应是不同的，但是这些反应具有重要的特点。每个聚合物的延长，要通过缩合反应将单体加到生长着的链状聚合物的末端，在此反应中，每增加一个亚单位就失去一分子水（图 2-19）。由单体逐步聚合成一条长链是生成巨大复杂分子的一种简单方法，因为整个聚合作用由同一组酶经过反复作用催化单体的加成。在某种意义上讲，这个过程就像工厂里机器反复操作，除了一个关键的方面以外。与某些多糖不同，大部分大分子都是从大量的彼此略为不同的单体合成的，如合成蛋白质的 20 种不同的氨基酸。聚合链并不是由单体随机组装而成；而是按特定次序（或序列）依次加入的，这对于生命是至关重要的。第 5 章和第 6 章详细论述了酶催化实现这一点的精确机制。

非共价键不仅决定了大分子的精细形状，而且决定了它与其他分子的结合

大分子中大部分共价键允许它们连接的原子旋转，所以聚合链具有很大的柔性。从理论上讲，这使得大分子在随机热能的影响下缠绕和旋转时具有无数种形状或者构型。然而，由

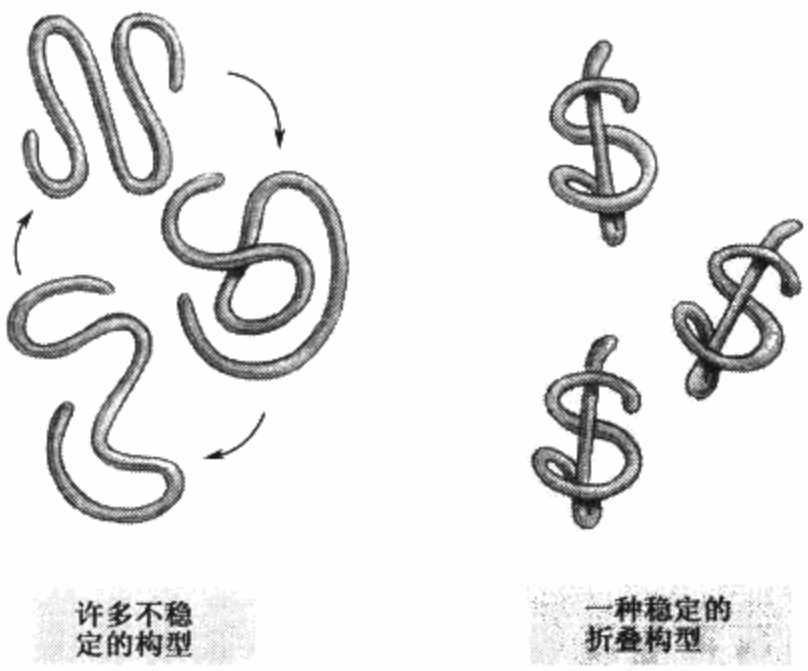


图 2-31　大部分蛋白质和许多 RNA 都折叠成唯一的一种稳定构型

如果维持这个稳定构型的非共价键受到破坏，那么分子就变成一条柔软多变的链，通常没有任何生物学价值

于在分子不同部位之间形成弱的非共价键，大部分大分子形状是高度受限的。如果形成足够多的非共价键，聚合链就会强烈倾向呈现某一特定的构型，它由链中单体的线性序列决定。事实上，细胞内发现的所有的蛋白质分子和许多小 RNA 分子都以这种方式紧密折叠成这种高度倾向的构型（图 2-31）。

生物分子中重要的 4 种非共价相互作用已在本章前面论述过，也可见专题 2-3。尽管它们中每个都很微弱，但是这些相互作用不仅共同作用使生物大分子折叠成独一无二的形状：在不同的两种分子之间，如果它们彼此紧密吻合，这些相互作用也会共同作用在这两种分子间形成一种强烈的相互吸引，就像手套戴到手上一样，这种形式的分子相互具有很大特异性。由于牢固结合所需的多点接触也使大分子能够从细胞内存在的大量其他类型分子中仅选出一种（通过结合）最合适的分子，而且，由于结合的强度有赖于形成的非共价键数目，所以几乎任何一种亲和相互作用都是可能的，一旦需要，就可以迅速解离。

这种类型的结合是所有生物催化剂发挥作用的分子基础，因而可以使蛋白质能够发挥酶的作用。非共价键也使得大分子能够用于构建更大的结构。在细胞内，大分子经常彼此结合起来形成巨大的复合体，因而组装出复杂的机器，拥有多种自由活动的部件，能够行使诸如 DNA 复制和蛋白质合成这样复杂的功能（图 2-32）。

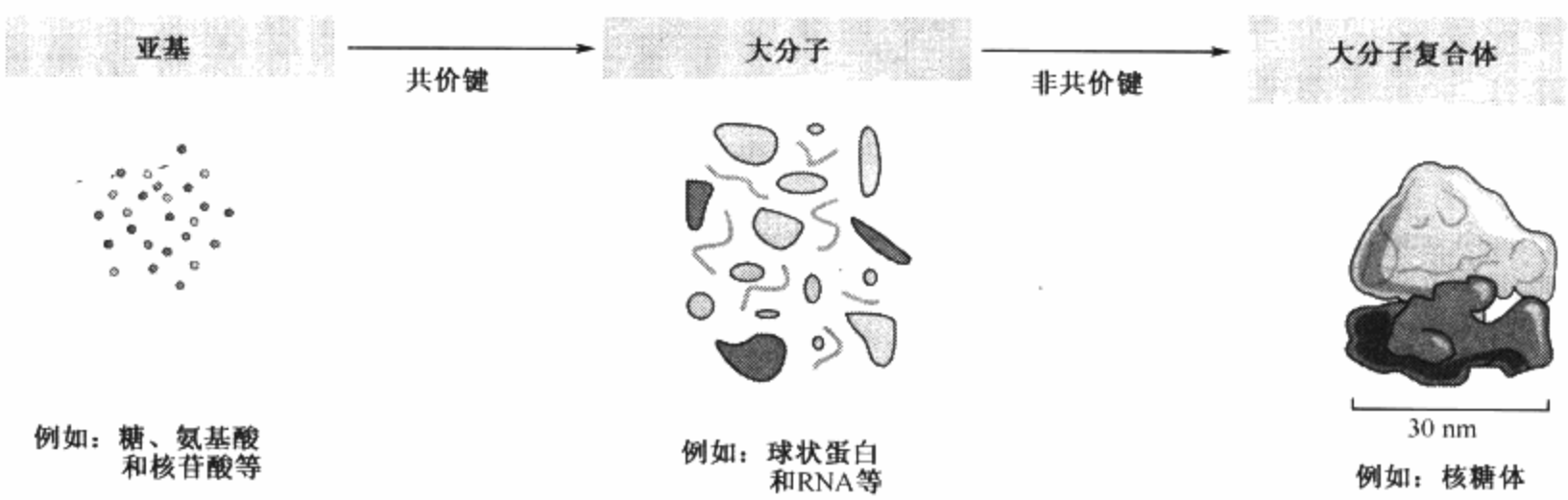


图 2-32　小分子、蛋白质和按近似比例画的核糖体

核糖体是细胞蛋白质生产机器的主要部分，每个核糖体都是约 90 个大分子（蛋白质和 RNA 分子）形成的复合体

小结

活生物体是一个自我调控、自我繁衍的化学系统。它们由一系列不同的有限数量的含碳小分子的组成，对于任何一种物种来讲，本质上都是一样的。这些分子中每一种都由一些小原子构成，这些原子通过共价键彼此相连呈现出某种精确的构型。

这些小分子主要是糖、脂肪酸、氨基酸和核苷酸。糖是细胞的主要化学能量来源，而且可以组装成多糖作为能量储备。脂肪酸对于能量储存也很重要，但它们最重要的作用是参与细胞膜的形成。由氨基酸组成的聚合物都是差别显著且功能多样的大分子，称为蛋白质。核苷酸在能量转移过程中发挥主要作用。核苷酸也是形成信息分子，RNA 和 DNA 的亚单位。

细胞的大部分干物质都是大分子，它们是氨基酸（蛋白质）或者核苷酸（DNA 和 RNA），按特定的顺序彼此共价连接起来形成的线性聚合物。蛋白质和许多 RNA 分子折叠成独特的构型，它由分子中亚单位的序列决定。这个折叠过程产生独特的表面，折叠过程也依赖于原子间由非共价力产生的大量弱相互作用。这些力有 4 种：离子键、氢键、范德华力和非极性基团间的由于疏水性而从水中排斥导致相互作用。同一类弱相互作用力控制其他分子

和大分子的特异性结合，使得生物分子间的各种协同能够产生细胞的结构和细胞内的化学。

催化作用和细胞利用能量

最重要的是，生物的一个性质使它们看起来几乎惊人地不同于无生命物质：生物在一个总是趋于更加混乱的宇宙中创造和维持有序性（图 2-33）。为了创造有序性，生物体内的细胞必须不断地进行化学反应。在某些这些反应中，有机小分子——氨基酸、糖、核苷酸以及脂，被分解或修饰以供给细胞所需的许多其他小分子。在其他反应中，这些小分子被用来构建非常多的蛋白质、核酸以及其他的大分子，这些大分子赋予生命系统所有的最有特色的性质。每个细胞可以被视作一个小型的化学工厂，每秒进行数百万的反应。

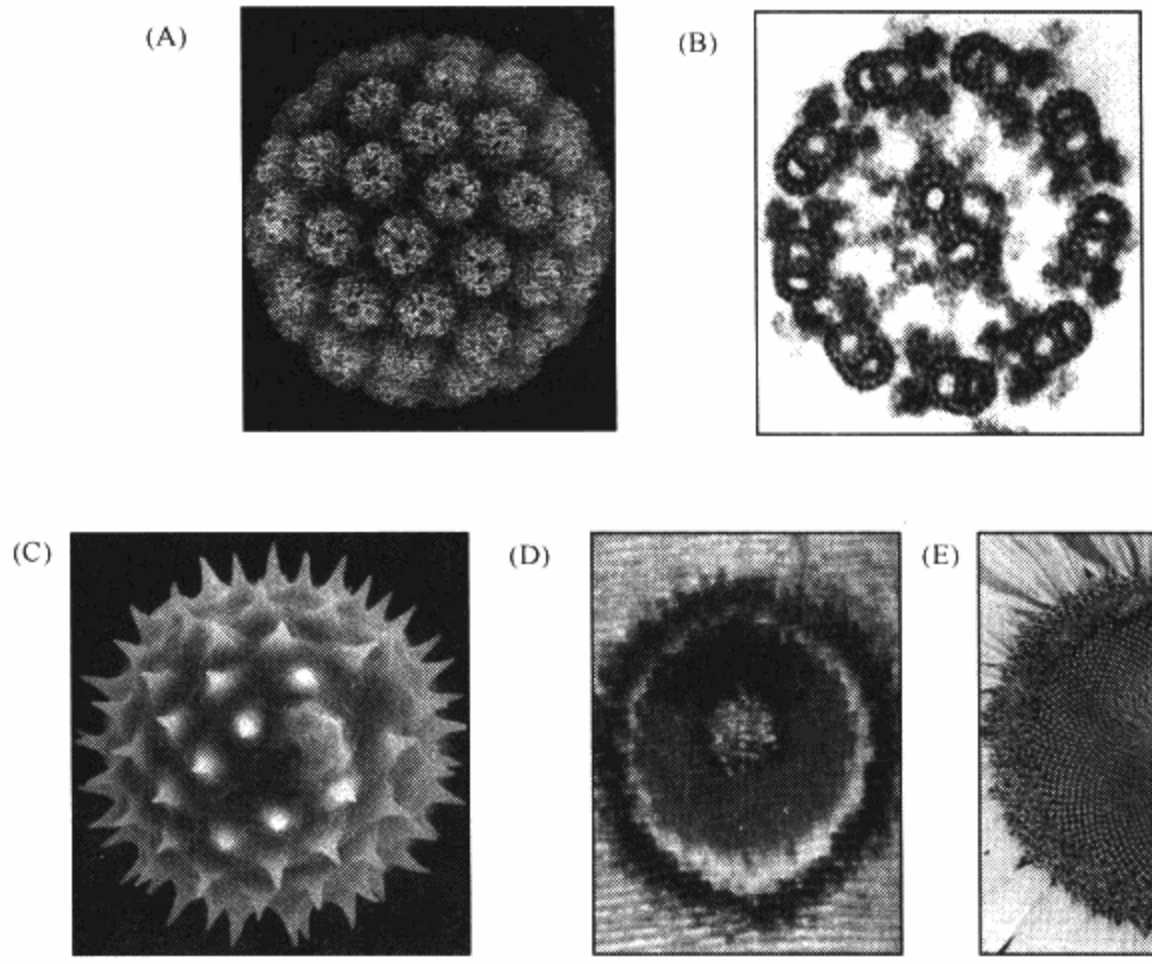


图 2-33　生物结构中的有序性

活生物体的各个组织水平都存在排列整齐的、华丽的、漂亮的空间图案。按从小到大的顺序：（A）病毒外壳的蛋白质分子；（B）精子尾部的微管以横截面看到的规则排列；（C）花粉颗粒（单个细胞）的表面轮廓；（D）蝴蝶翅膀的闭合显示由翅瓣形成的图案，每个翅瓣是由一个细胞形成的；（E）向日葵中螺旋排列的种子，它们由数百万个细胞组成。[（A）经R. A. Grant 和 J. M. Hogle 许可使用；（B）经 Lewis Tilney 许可使用；（C）经 colin MacFarlane 和 Chris Jeffree 许可使用；（D）和（E）经 Kjell B. Sandrde 许可使用]

细胞代谢是由酶组织的

细胞所执行的化学反应通常只在高于细胞内的温度发生。由于这个原因，每个反应需要化学反应性的特异增强。该需要是重要的，因为它使得每个反应受到细胞的控制。这种控制是通过称为酶的特化蛋白质执行的，每个酶只加速或催化特定分子可能进行的许多种反应中的一种。酶催化的反应通常串联在一起，以便一个反应的产物称为另一个反应的起始材料或底物（图 2-34）。这些长的线性反应途径反过来彼此连接，形成互联反应的网络，使得细胞能够生存、生长以及繁殖（图 2-35）。

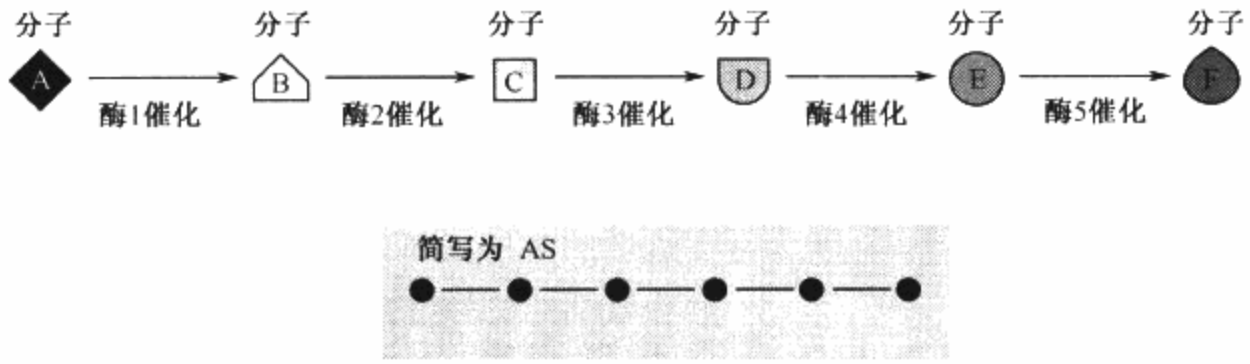


图 2-34　一组酶促反应如何形成一条代谢途径

每种酶催化一个特定的反应，从催化的过程中酶保持不变。在这个例子中，依次作用的一组酶将分子 A 转变为分子 F，从而形成一条代谢途径

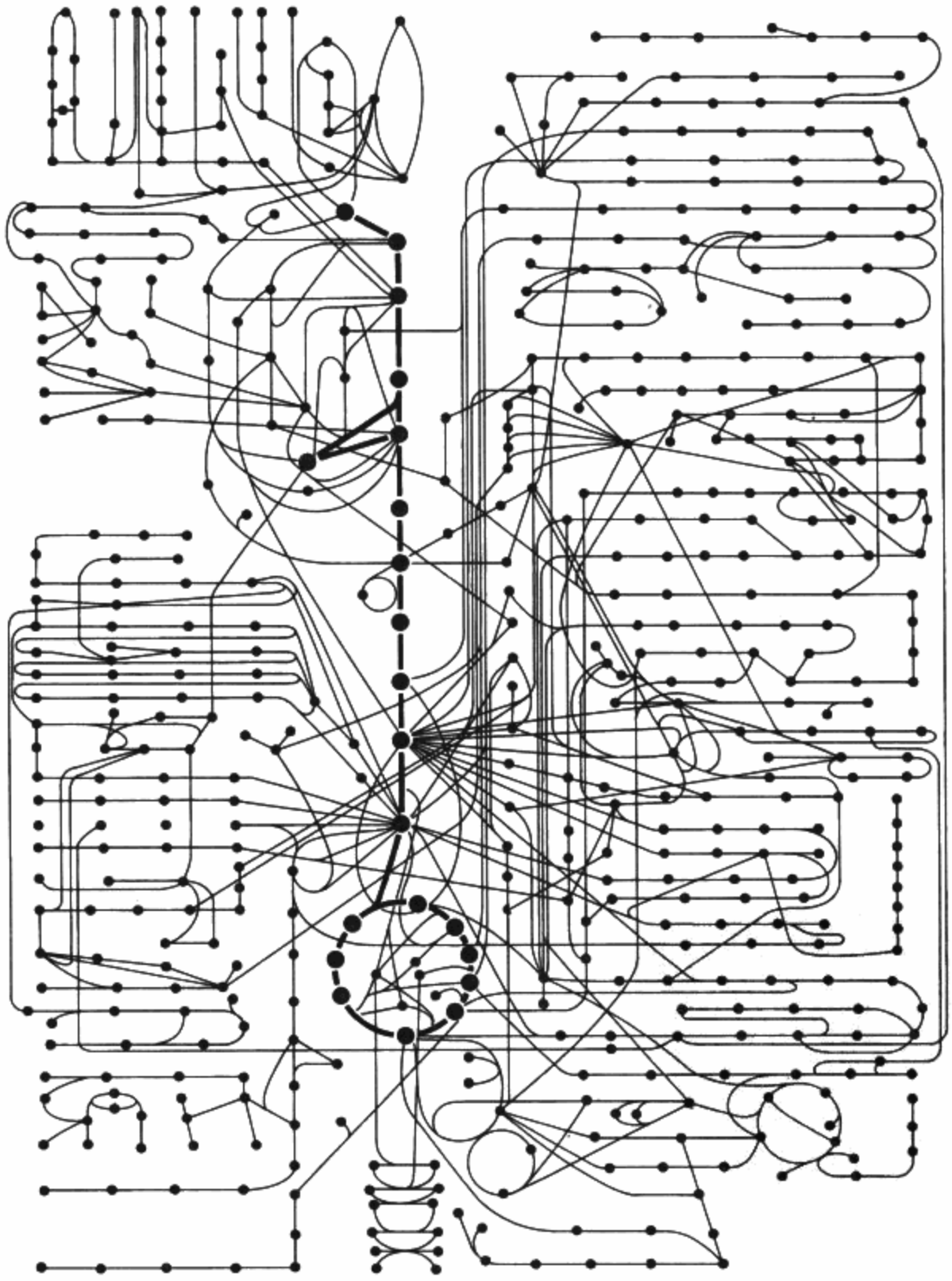


图 2-35　某些代谢途径以及它们自己在一个典型细胞内的互相联系

图中标出大约 500 个共同的代谢反应，一个代谢途径中的每个分子用一个实心圈代表，如图 2-34 中黄色盒子所示

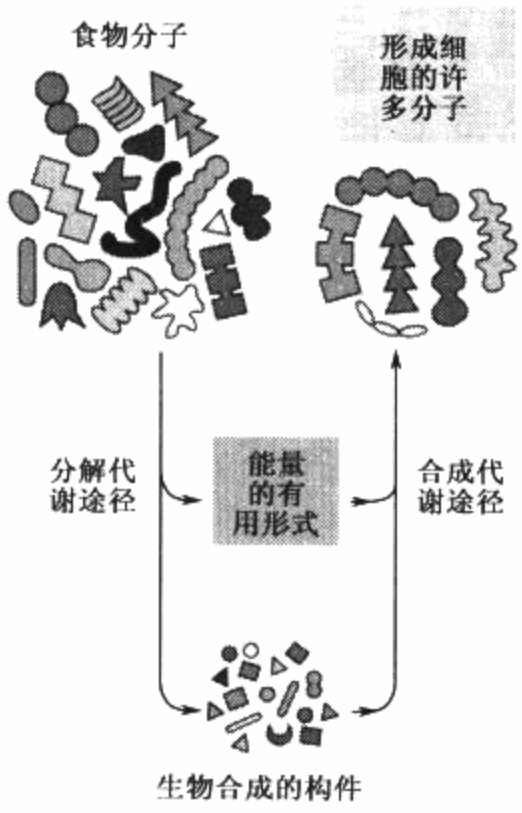


图 2-36　新陈代谢中分解代谢和合成代谢之间关系的图解

如图中所示，因为食物分子的化学键中储存的能量大部分以热的形式释放，所以能量全来源于分解代谢的任何生物所需的食物的质量大大多于合成代谢所生成的分子的质量

细胞内发生两种方向相反的反应：①分解代谢途径将食品分解成更小的分子，从而产生对细胞有用的能量形式和细胞所需的作为原材料的某些小分子；②合成代谢或生物合成途径利用分解代谢所产生的能量促进形成细胞的许多其他分子的合成。这两组反应一起构成细胞的**新陈代谢**（metabolism）（图 2-36）。

细胞新陈代谢的许多细节是生物化学的传统主题，在这我们并不关注这些细节。但是，细胞从环境中获得能量并利用能量创造有序性所采用的普遍原则对于细胞生物学是重要的。我们先介绍为何能量的不断输入对于维持生物体是必需的。

细胞释放的热能使得生物有序性成为可能

事物普遍趋于变得混乱可以用物理基本定律——热力学第二定律来表示，热力学第二定律是说在宇宙中或任何独立的系统（完全与宇宙的其余部分隔离的物质总和）中混乱程度只能增加。热力学第二定律对所有生物具有如此深远的影响，以至于该定律值得以几种方式重新介绍。

例如，我们可以用概率的方式介绍第二定律，我们说系统将自发地向那些具有最大概率的排列改变。例如，如果我们考虑一盒 100 个硬币朝上，一系列扰乱盒子的偶然事件将趋于将排列移向 50 个头和 50 个尾的混合。原因很简单：存在很多种单个硬币的可能排列能实现 50-50 结果，但只有一种可能排列使得所有的硬币方向朝上。因为 50-50 的混合是最可能的，所以我们说它更“混乱”。因为相同的原因，通常在没有有意的努力的情况下一个人的生活空间将变得日益混乱：混乱的运动是一种自发过程，需要定期的努力来逆转（图 2-37）。

一个系统中的混乱的量可以定量。我们用来测量这种混乱的量称为系统的**熵**（entropy）：混乱越大，熵则越大。因此，表示热力学第二定律的第三种方式是系统自发地向更大熵的排列改变。

活细胞经过存活，生长并形成复杂有机体的过程是在不断产生有序性的，因此看起来似乎违背了热力学第二定律。这怎么可能呢？答案在于一个细胞并不是一个隔绝的系统：它从环境中吸收能量，这些能量的存在形式，或者是食物，或者是太阳光子（更甚至于像化合细菌那样仅来自于无机分子），然后细胞利用这种能量在自身内部建立有序性。在此化学反应过程中，细胞所利用的能量一部分转化为热量，这些热量被释放到细胞所处环境中，并增加了环境的无序性，因此正如物理定律规定的那样，总熵（细胞与环境的总熵）增加了。

为了理解控制这些能量转换的规则，让我们设想一个细胞处于代表宇宙其余物质组成的“海洋”中。当这个细胞存活并生长时，它在内部产生有序性。但是当它合成分子并将它们组装到细胞结构中去时，释放了热能。热是以最无序的形式存在的能量——随机振动的分子。当此细胞将热释放到环境中时，它增加了分子运动的频率（热运动），因此增加了环境的随机度或者无序度。由于细胞内有序度的增加不足以抵消环境中有序度的减少（熵增加），所以这个过程符合热力学第二定律（图 2-38）。

图 2-37　自发趋向混乱的日常图解

逆转趋向于混乱的趋势需要有意的努力和努力的输入：它不是自发的。事实上，从热力学第二定律我们可以确定所需的人的干预将释放足够的能量到环境中，这些能量不足以补偿重新布置该房间内物品所耗的能量

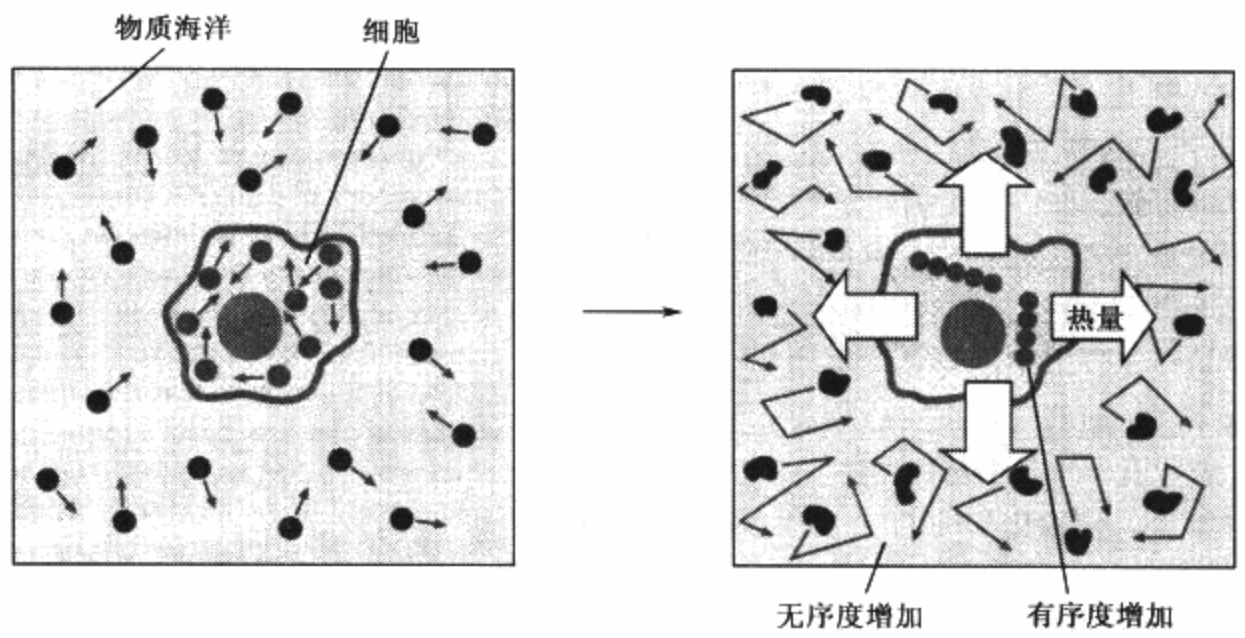


图 2-38　活细胞的一种简单的热动力分析

左图表示细胞内部和宇宙其余空间中（物质海洋）的分子处于相对紊乱的状态。右图中细胞通过一个增加其内部分子有序度的反应，从食物分子中吸收能量，并释放出热量。由于热增加了细胞周围环境中的无序度（用锯齿状的箭头和扭曲的分子表示，表明热增加了分子运动），所以在细胞生长和分裂的过程中遵守了热力学第二定律：该定律认为宇宙中无序度的总量必须不断增加（更详细的讨论见专题 2-7）

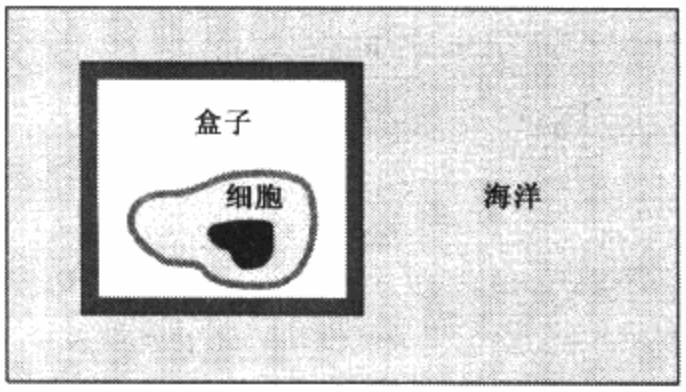
细胞释放的热量从哪里来呢？这里我们遇到另外一个重要的热力学定律。热力学第一定律规定能量可以从一种形式转变为另一种形式，但是既不能产生也不能消亡。图 2-39 列举了某些形式的能量。由于细胞内发生化学反应，以不同形式存在的能量的数量将发生改变，但第一定律告诉我们，能量的总量必须保持不变。例如，某一个动物细胞吸收食物，并将存在于这些食物分子内原子间的化学键能转变成分子的随机热运动（热量）。如果这些细胞内的反

专题 2-7　自由能和生物反应

自由能对细胞的重要性

生命之所以成为可能，是由于每个细胞中发生的化学反应相互作用，形成复杂的网络。当我们考虑组成这个网络的代谢途径时，可能会怀疑细胞能进化产生一种酶，催化任何它需要的反应。但事实并不是这样。尽管酶是强有力的催化剂，但是它们仅仅能够催化那些热动力学可能的反应；其他的反应之所以能在细胞内进行，仅仅是因为它们与推动自身进行的非常有利的反应偶联起来。一个反应是否能够自发进行，还是需要与其他反应偶联起来，这个问题对于细胞生物学是很重要的。借用自由能这个量我们可以得到答案。即在一系列反应过程中，总自由能变化决定整个序列反应能否发生。在本专题中，我们将阐述一些基本理论来自化学和物理学的一个特殊分支，称为热动力学，以帮助理解什么是自由能以及为什么对细胞如此重要。

化学键发生改变所释放的自由能转变成热能



一个封闭的系统定义为一个分子的集合，它不和宇宙其余部分交换物质（如上图所示的“盒子里的细胞”）。任何这种系统都含有总能量为 E 的分子。这个量将以不同的形式分布：一些作为分子的平移能，一些作为分子的振动和旋转的能量，但是大部分是组成分子的单个原子之间的键能。假设系统内发生反应，热力学第一定律限定了什么类型的反应能发生，它规定任何反应过程中，宇宙能量的总量保持不变。例如，假设盒子中某个位置发生了 A → B 的反应，并且释放了大量的化学键能。这些能量最初将增加系统内部分子运动的频率（平移的、振动的和旋转的），这也相当于增加了系统的温度。然而，这些增加的运动将很快转移到系统外，它是通过一系列的分子碰撞，首先增加了盒壁的温度，然后增加盒子外的温度［在我们的例子中用海洋（sea）代表盒子外］。结果系统恢复到最初的温度，而此时盒子内部释放的所有化学键能也全部转变成热能，并转移到盒子外面的环境中。根据第一定律，盒内能量的变化（$\Delta E_{盒子}$ 我们将用 ΔE 表示）必须与转移的热能（我们用 h 代替）在数量上相等，而符号相反，即 $\Delta E = -h$。因此，当热能从系统释放时盒内的总能量（E）减少。反应中 E 也可以因为盒子外部做的功而改变。例如，假设在某个反应中，盒子的体积（ΔV）发生了较小的改变。由于盒子壁必须逆着外界环境的恒定压力（P）向外推动以扩大体积。这个过程对外界环境做了功，需要能量。所用的能量为 P（ΔV），根据热力学第一定律，盒内必须减少相同的量。在大多数反应中，化学键能都转变成功和热。焓（H）是一个混合的方程，包含两者（$H = E + PV$）。严格意义上讲，反应过程中是焓变而不是能量的变化与转移到外部环境中的热能相等。在 H 降低的反应中，向环境释放了能量，因此说这类反应是“放热的”；而在 H 增加的反应中，从环境中吸收了能量，因此说这类反应是“吸热的”。因而，$-h = \Delta H$。然而，在大部分生物反应中，体积改变可以忽略不计，所以得到更好的近似：

$$-h = \Delta H \cong \Delta E$$

热动力学第二定律

假设有一个容器里装了 1000 枚硬币，且全都正面朝上，如果用力摇动这个容器，让硬币做类似所有分子因频繁与其他分子碰撞所做的那种随机运动。那么结果将会是大约一半的硬币将正面朝下。发生重新定向的原因在于：使最初处于有序状态的硬币在运动后重新恢复原状（每个硬币正面向上）的方式只有一种，然而却有许多不同的方式（大约 10^{298}）使其处于 50-50 状态的方式比任何其他状态的方式都更多。每一种状态都有一个发生的概率，它与现实这个状态存在的方式的数目呈正比。热力学第二定律表述为：系统将自发地从较低可能性的状态转变为较高可能性的状态。由于较低可能性的状态比较高可能性状态更加有序，所以第二定律可以重新表述为：宇宙不断发生变化以达到更加无序的状态。

熵，*S*

第二定律（而不是第一定律）可以用来预测特定反应进行的方向。但是为了使其能用于这个目的，就必须有一个方便的测量方式，借以得到无序状态的程度或者概率大小。熵就是这样一种测量方式，它是概率的对数方程

$$\Delta S = R\ln\ p_B/p_A$$

这样当其中 p_A 和 p_B 是状态 A 和 B 出现的概率时，R 是气体常数（2 cal deg^{-1} mol^{-1}），ΔS 用熵单位（eu）衡量。在前面 1000 个硬币的例子中，所有正面向上（状态 A）和一半向上一半向下（状态 B）的概率比值与实现这两种状态的方式数目的比值是相同的。可以计算一下：$p_A = 1$ 和 $p_B = 1000!$ $(500! \times 500!) = 10^{298}$。因此，当用力摇动容器，使得正面向上和向下的硬币各占一半时，硬币重新定向的熵度是 R ln（10^{298}），或者大约每摩尔容器（6×10^{23} 个容器）1370eu。我们看到，因为在状态 A 向状态 B 过渡（p_B/p_A）中上面定义的 ΔS 为正，所以具有大的 S 增加的反应（即 $\Delta S>0$）在能量上有利，更倾向于自发进行。

正如第 2 章讨论的，热能引起分子的随机运动。由于热能从一个封闭的系统转移到周围环境中，增加了外界分子不同排列的数目，所以它增加了这些分子的熵。可以证明，释放一定数量的热能，在低温条件下比在高温条件下可以产生更大的无序性效果；此外也可以证明环境的 ΔS 值（上面定义的 $\Delta S_{海洋}$）与转移到环境中的热能完全相等，$\Delta S_{海洋}$ 等于系统（h）除以绝对温度（T）：

$$\Delta S_{海洋} = h/T$$

吉布斯自由能，*G*

当研究一个封闭的生物系统时，更愿意用一种简单的方式预测给定的反应能否在系统内自发进行。我们已经看到关键问题在于宇宙的熵变在反应发生时是正的还是负的。在假定的理想系统（盒内的细胞）内，宇宙的熵变由两部分组成，封闭在盒子里面的系统的熵变以及周围的“海洋”的熵变，在做出任何预测前必须将二者加合起来。例如，一个反应可能吸收热量，因而减少了海洋的熵（$\Delta S_{海洋} < 0$），同时增加了盒内的无序性程度（$\Delta S_{盒子} > 0$），这样全部的 $\Delta S_{宇宙} = \Delta S_{海洋} + \Delta S_{盒子}$ 大于零。这种情况下，反应将自发进行，即使反应过程中海洋将热传递给盒子。这类反应的一个实例是氯化钠在盛有水的烧杯（“盒子”）中溶解，这是一个自发进行的过程，即使当盐溶解到溶液中时水温降低了。

化学家发现，定义一些新的“混合方程”，对于描述系统的结合在一起的物理性质很有用。可以结合在一起的性质包括温度（T）、压力（P）、体积（V）、能量（E）和熵（S）。焓（H）则是混合方程之一。但是，到目前为止对于生物学家最有用的混合方程是吉布斯自由能，G。它充当一个计算装置，用来推理出盒子内发生化学反应引起的宇宙熵变的大小，同时避免分开考虑海洋熵变的麻烦。G 的定义是

$$G = H - TS$$

式中，对于体积为 V 的盒子，H 是上面所讲的焓（$E+PV$），T 是绝对温度，S 是熵。这些量中每个都仅适用于盒内反应。在盒内进行的反应中，自由能变化（产物的 G 减去起始物的 G）用 ΔG 表示，正如我们将要证明的那样，它是反应发生时所产生的无序的数量的直接衡量。

恒温时反应过程中自由能的变化（ΔG）等于 $\Delta H - T\Delta S$。联系到 $\Delta H = -h$，即以海洋吸收的热量，我们现在有两个方程：

$$-\Delta G = -\Delta H + T\Delta S$$
$$-\Delta G = h + T\Delta S$$

所以

$$-\Delta G/T = h/T + \Delta S$$

但是 h/T 等于海洋的熵变（$\Delta S_{海洋}$），上面方程中 ΔS 即为 $\Delta S_{盒子}$。因此

$$-\Delta G/T = \Delta S_{海洋} + \Delta S_{盒子} = \Delta S_{宇宙}$$

我们得出结论：自由能变化是宇宙熵变的直接衡量方式。一个反应将沿着导致自由能变化小于零的方向进行，因为在这种情况下，当反应发生时，宇宙的熵变为正。

对于一系列涉及许多不同分子的偶联反应，总自由能变化可以这样简单地计算为：将反应后所有不同分子的自由能加和起来，然后与反应前这些分子的自由能之和进行比较；对于常见的物质，所需的自由能变化可以从发表的表格中查到。通过这种方式，可以预测反应进行的方向，从而可以很容易地检验任何假设机制的可行性。例如，从观察到的跨线粒体内膜的电化学质子梯度数量级和 ATP 在线粒体水解的 ΔG 当中，可以肯定 ATP 合成酶每合成 1 分子 ATP 需要多于 1 个质子通过线粒体膜。

一个反应的 ΔG 值可以直接衡量该反应离平衡有多远。细胞内 ATP 水解反应巨大的负 ΔG 仅仅反映出细胞将 ATP 水解反应维持在离平衡状态多达 10 次方的数量级上。一个反应如果达到平衡，$\Delta G = 0$，反应就会以完全相同的速率向正反应和逆反应方向进行。对于 ATP 水解反应，当大量的 ATP 被水解时，反应到达了平衡，这种情况发生在死亡的细胞中。

应是为了使宇宙整体变得更加无序，那么这种化学能向热能的转化就是必需的，正如第二定律规定的那样。

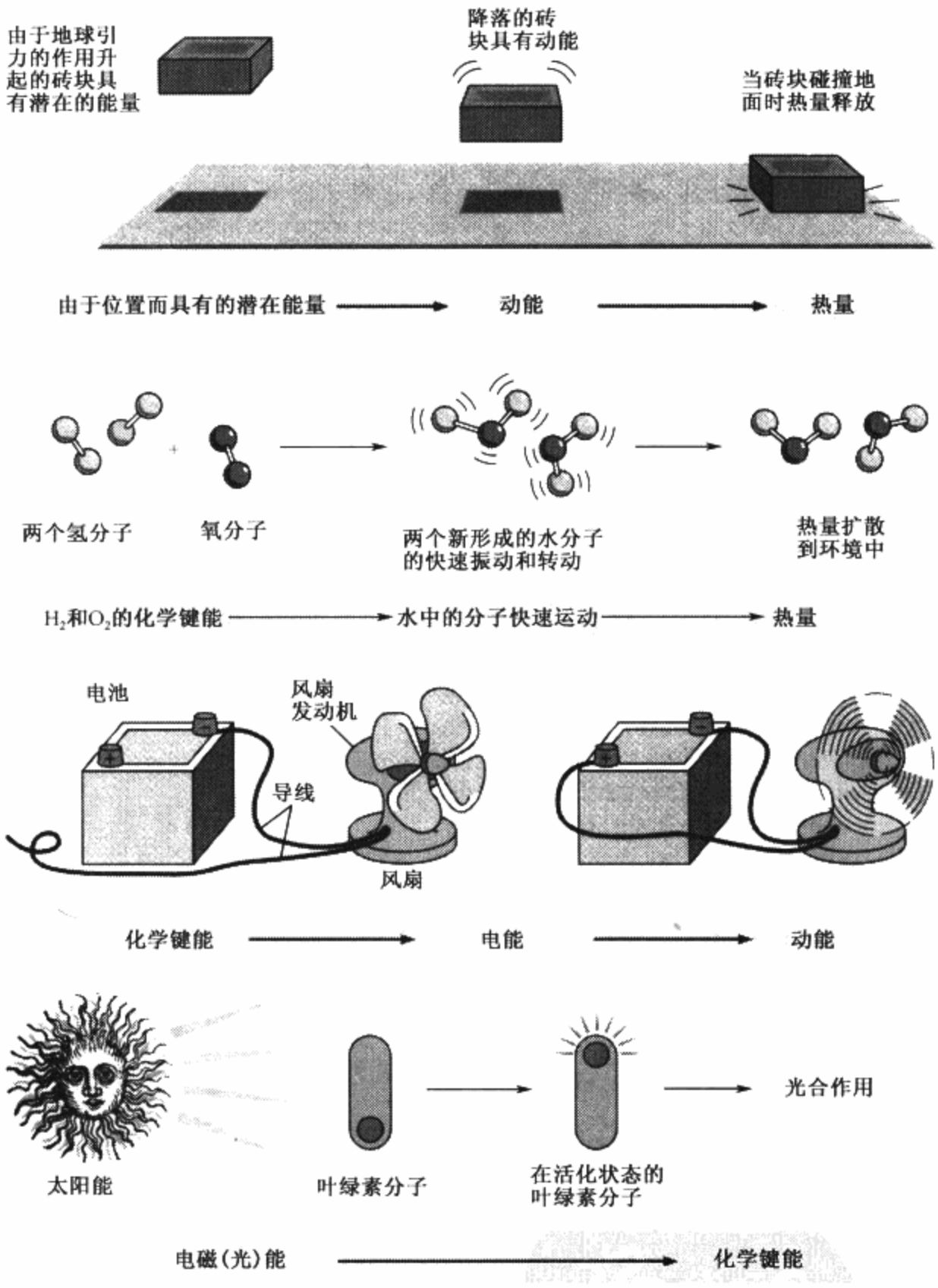


图 2-39　一些不同能量形式之间的相互转化

原则上所有的能量形式都可以相互转化。在所有这些过程中，能量的总量保持不变。因此，例如，从第一个例子砖块的高度和重量中，我们可以推测出它落地时有多少热释放。在第二个例子中，注意当水分子形成时，大量的化学分子键能最初被转化为两个新生成的水分子内非常迅速的热运动，但是与其他分子间的碰撞几乎瞬间就将这种动能均匀地分散到环境中（热转换），从而使得新生成的分子与其他分子一样

细胞并不能从它释放的热量中获益，除非细胞内产生热反应直接与分子有序性的产生过程相连。正是产热与有序度的增加之间紧密偶联，使得细胞的新陈代谢显著有别于燃料在火中消耗性的燃烧。在本章的后部分，我们将阐明这种偶联是怎样进行的。到此，我们完全有

理由认识到，如果细胞能够在倾向于混乱的环境中建立并维持有序状态，那么必须使食物分子的燃烧同生物学有序性的产生直接联系起来。

光合生物利用阳光合成有机分子

所有动物的生存都依赖与储存在有机分子化学键中的能量，这些有机分子由其他生物合成，并被动物采食。这些分子也提供给动物所需的原子以构建新的活性物质。有些动物靠捕食其他动物获得食物，但处在动物食物链最底层的是采食植物的动物。这些植物反过来，从阳光中捕获能量结果，所以动物细胞利用的能量最终都来源于阳光。

光能通过植物和光合细菌中的**光合作用**进入生命世界，光合作用使得阳光中的电磁能转化为细胞内的化学键能。植物能从无机物中获得它们所需的所有原子：从大气二氧化碳中得到碳，从水中得到氢和氧，从土壤里的氨及硝酸盐中得到氮，还能从土壤里的无机盐中得到需求量较少的其他元素。植物利用光获得的能量将这些原子组装成糖、氨基酸和脂肪酸。这些小分子反过来进一步转变成组成植物的蛋白质、核酸、多糖以及脂类物质。一旦后来这些植物被动物采食，所有这些物质都成为动物的食物分子。

光合作用发生在两个阶段（图 2-40）。在第一阶段，光能被捕获并作为化学键能瞬间储存于特殊的小分子中，这些小分子作为能量和化学基团的载体发挥作用（我们后面将讨论这些活化的分子载体）。光使水分子裂解释放出分子氧，作为第一阶段的无用产物被释放。

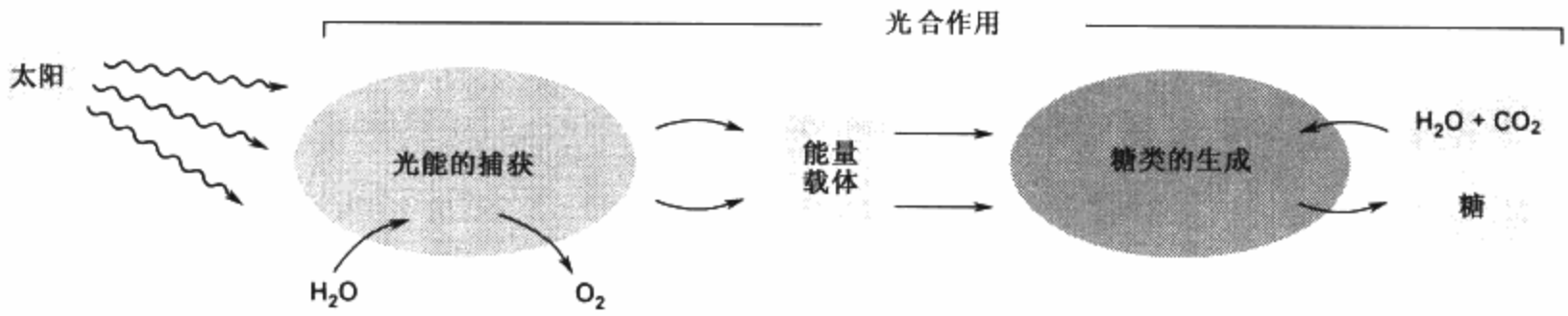


图 2-40　光合作用

光合作用的两个阶段。第一阶段产生的能量载体是后面将要讨论的 2 个分子——ATP 和 NADPH

在第二阶段，能量载体分子用于促进碳的固定过程。在这个过程中，利用 CO_2 提供一个气体和水分子中生成糖，因此对于植物自身以及任何采食植物的动物来讲都提供了一种有用的途径以储存化学键能和物质。我们将在第 14 章中描述这两个光合作用阶段的精妙机制。

整个光合作用过程的净结果，就绿色植物而言，可以简单地用下面的方程概括：

$$光能 + CO_2 + H_2O \longrightarrow 糖 + O_2 + 热能$$

生成的糖接着用于提供化学键能，或者作为生成植物细胞必需的各类大小有机分子的原料。

细胞通过氧化有机分子获取能量

所有的动物和植物细胞都由储存在有机分子内的化学键能提供动力，无论这些有机分子是植物所为食物而光合生成的糖，还是动物采食的大小分子。为了利用能量生存，生长并繁殖，有机体必须以某种可利用的形式获取它。在动物和植物中通过逐步氧化或有控制的燃烧过程从食物分子中获取能量。

地球大气层含有大量的氧气，在氧气存在的条件下，能量上最稳定的碳是 CO_2，最稳定的氢是 H_2O，因此细胞可以将糖或其他有机分子的碳原子、氢原子与氧结合，分别生成 CO_2 或 H_2O，从而获取能量，这个过程称为**呼吸作用**（respiration）。

光合作用与呼吸作用是互补的两个过程（图 2-41）。这意味着动物与植物之间的能量交易并非全是单向的。植物、动物以及微生物在这个星球上共存了这么久，以至于它们中许多已成为其他成员的环境中重要组成部分。几乎所有的有机体都将光合作用释放的氧在有机分子的燃烧中消耗掉。并且一片绿叶中前一天通过光合作用固定到有机分子中的部分二氧化碳分子，第二天由于动物的呼吸作用（或者是降解死亡有机物质的真菌和细菌的呼吸作用）就被释放到空气中。因此从整体上看，碳的利用形成了一个涉及生物圈（气球上所有生命的有机体）的巨大循环，它跨越了单个有机物间的界限（图 2-42）。同样，氮原子、磷原子和硫原子也在涉及植物、动物、真菌和细菌的循环中，在有生命和无生命的世界之间流动。

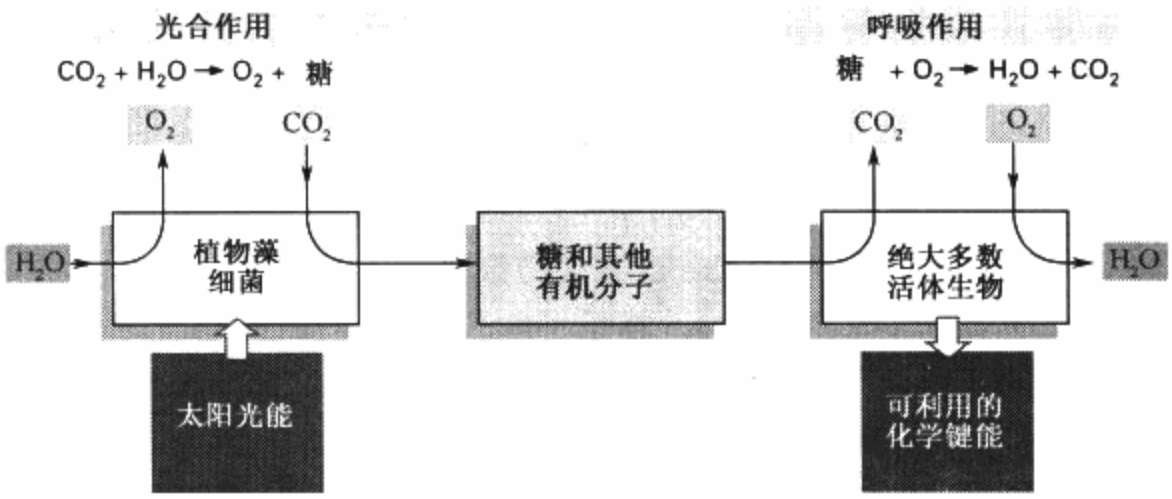


图 2-41　生命世界里光合作用与呼吸作用是互补的过程

光合作用利用光能生成糖和其他有机分子，这些有机分子反过来又作为其他有机体的食物。许多种有机体可以进行呼吸作用，这个过程利用氧气从碳原子生成二氧化碳，而这个碳原子恰恰是光合作用过程作为 CO_2 吸收并转变成糖的同一个碳原子。通过这个过程，进行呼吸作用的有机体获得它们生存必需的化学键能。人们认为地球上第一个细胞既不能进行光合作用，也不能进行呼吸作用（第 14 章讨论），但是地球上光合作用必须先于呼吸作用发生。因为存在强有力的证据表明，释放足够量的氧气以形成富含这种的气体的空气需要进行数十亿年时间的光合作用（地球大气层目前含有 20%的氧气）

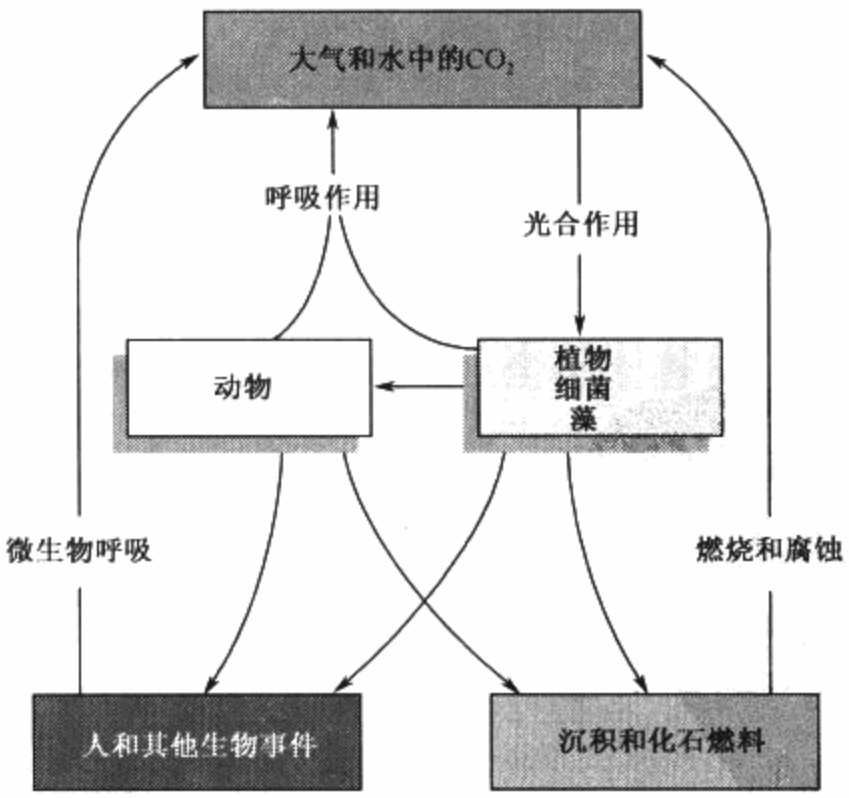


图 2-42　碳循环

单个碳原子通过植物细菌和海藻的光合作用整合到生命世界里的有机分子中。它们经由循环途径进入动物、微生物以及土壤和海洋中的有机物质内。当有机分子被细胞氧化或者被人类当作化石燃料烧掉后，CO_2 就重新进入空气中

氧化与还原涉及电子转移

细胞并不会像有机物在火中燃烧那样，一步氧化有机分子，通过利用酶催化剂，分子在新陈代谢中经历了大量反应，其中很少涉及直接的加氧反应。在我们考察这些反应以及它们的作用之前，我们需要先讨论一下氧化反应过程的含义。

氧化，就上述意义而言，并不仅仅是指氧原子的加成。相反，通常它适用于任何发生在原子之间的电子转移反应，在此意义上氧化是指电子的转移；而**还原**与氧化相反，是指电子加成。因此，Fe^{2+} 一旦失去一个电子变成 Fe^{3+}，它就被氧化，而一个氯原子一旦获得一个电子变成 Cl^-，它就被还原。由于一个化学反应中电子数目是保守的（既不丢失也不获得），氧化和还原总是同时发生；也就是说，某个反应中如果一个分子获得一个电子（还原），另一个分子就失去这个电子（氧化）。例如，当糖分子被氧化成 CO_2 和 H_2O 的时候，生成 H_2O 的氧分子获得电子，因而说 O_2 被还原了。

共价相连的原子之间，即便只有部分电子转移，“氧化”和“还原”这两个词同样适用（图 2-43）。例如，如果一个碳原子与另一个强亲电子的原子，如氧、氯、硫，共价结合，它就释放出比共用电子更多的电子，形成一个极性共价键：现在碳原子核的正电荷略多于它的电子所带的负电荷，因此原子带部分正电荷，称为被氧化了。相反，C—H 连接中的碳原子具有比它共有电子更多的电子，因此说它被还原了（图 2-43）。

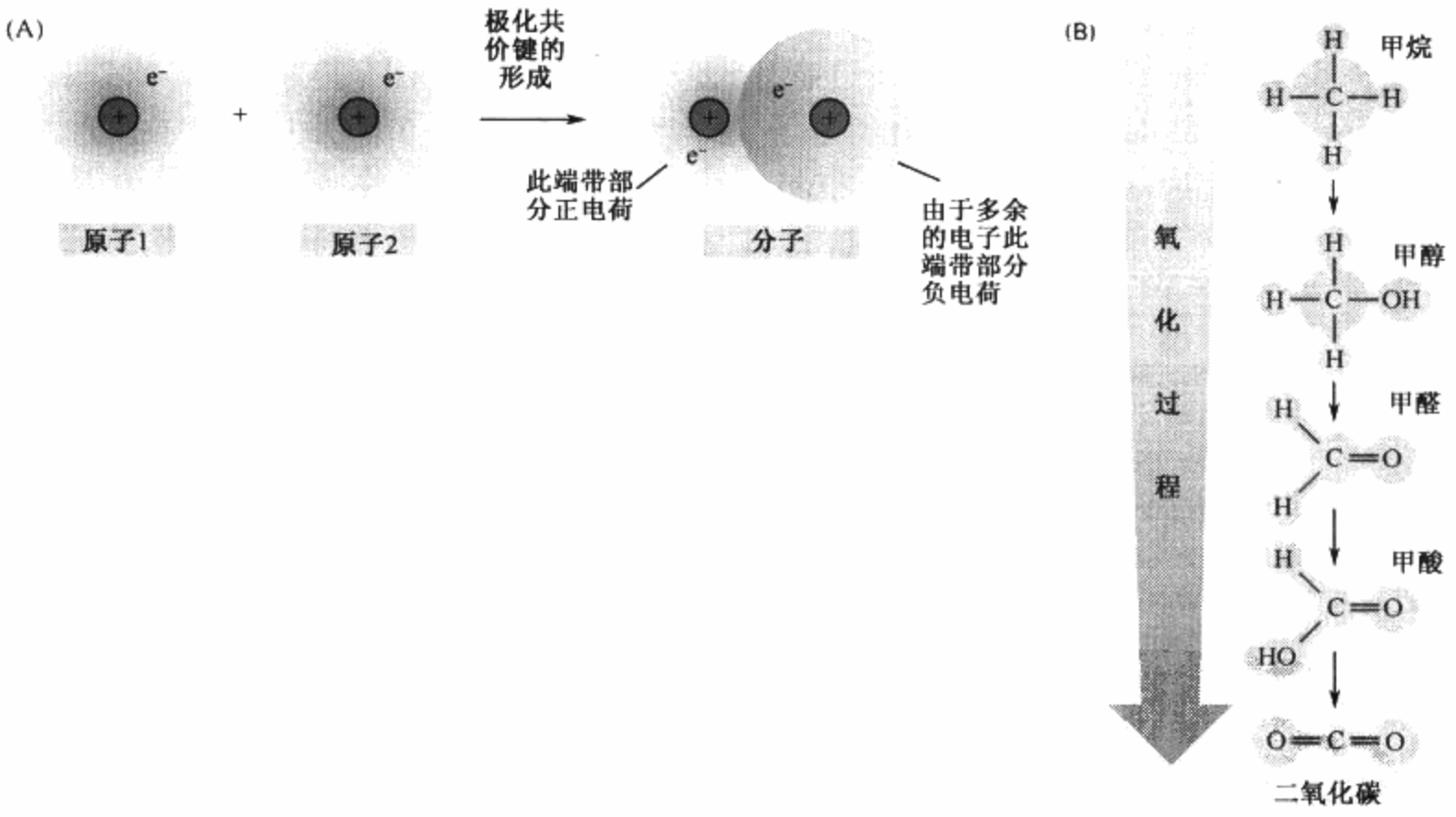


图 2-43　氧化和还原

两个原子形成一个极性共价键，具有更多共享电子末端的原子称为被还原，而其他获得较少共享电子的原子称为被氧化。由于原子核的正电荷不及环绕它的电子所带的负电荷，因而被还原的原子带部分负电荷（S^-）；相反，被氧化的原子带部分正电荷（S^+）。(B) 甲烷中的单个碳原子，通过用氧原子逐步替换共价结合的氢原子，可以转变成二氧化碳。每一步中，电子从碳原子上转移（用蓝色阴影表示），这样碳原子就不断变得更加氧化。在细胞内存在的条件下，每一步在能量上都是有利的

细胞内的分子获得一个电子（e^-）时，它通常也同时获得一个质子（水中游离子存在的

质子），这个例子的净效应是向分子中增加一个氢原子

$$A+e^-+H^+\longrightarrow AH$$

尽管只涉及一个质子加上一个电子（而不是仅仅一个电子），这种氢化反应属于还原反应；相反，脱氢反应属于氧化反应。一个有机分子是被氧化还是被还原，很容易区别。如果C—H键数目增加了，就发生还原反应；然而如果减少，就发生氧化反应［图2-43（B）］。

细胞经过一系列能够获得能量的反应，利用酶的催化作用，一点一点地氧化有机分子。现在我们需要解释一下酶是怎样发挥作用的，以及它们的一些限制条件。

酶降低了阻遏化学反应的障碍

看看下面的反应：纸＋氧气⟶烟尘＋灰烬＋热量＋二氧化碳＋水

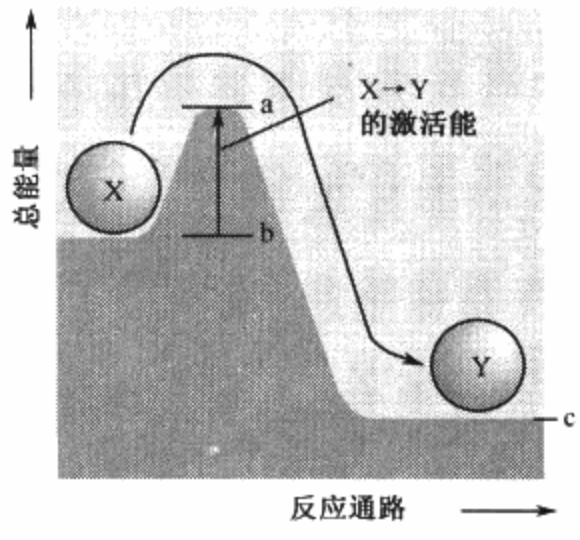


图2-44　活化能的重要原则

复合物X处于相对稳定的状态，将其转变为复合物Y需要能量，即便Y处于一个比X更低的能量水平上。因此，这种转变将不可能发生，除非复合物X可以从周围环境中获得足够的活化能（能量a减去能量b），从而进行反应，将它转变成复合物Y。这种能量也可以通过同其他分子异常的能量碰撞而得到。对于逆反应Y→X，活化能会更多（能量a减去能量c）；因此，这个反应发生可能性更小。活化能通常为正；然而，注意对于能量上有利的反应X→Y，总能量的变化是负数，即能量c减去能量b

纸很容易燃烧，并释放热能；水蒸气和二氧化碳气体到空中，但是烟尘和灰烬绝不会自发地从加热空气中获得这些组分，重新组装成纸张。纸燃烧以后，它的化学能作为热能散失掉，不会从宇宙消失，因为能量既不能创造出来，也不会消失，而是不可逆转地分散到分子混乱的随机热运动中，与此同时，组成纸张的原子和分子分散并无序化。从热动力学水平来说，就是发生自由能，即可用来做功并推动化学反应的能量的丢失。这种能量的丢失反映出储存在纸内的分子和能量的有序性丢失。我们很快将更详细地讨论自由能。但通常的原则是不言自明的：化学反应紧紧沿着自由能丢失的方向进行，换句话说，任何反应自发进行的方向都是“顺坡而下”的，在这个意义上讲，一个“顺坡而下”的反应通常称为能量上有利。

尽管通常条件下，碳在能量上最有利的形式是CO_2，氢的形式是H_2O，但是一个具有生命的有机体并不会消失于一缕青烟，你手中的书也不会自燃。这是因为活生物和书本中的分子处于相对稳定的状态。如果不额外施加能量，它们很难进入一个更低能量的状态。换句话讲，一个分子需要**活化能**（activation energy）——促进越过能量障碍的一种推动力，在它能进行化学反应之前，这种化学反应可使其处于更稳定的状态（图2-44）。在这个书燃烧的例子中，活化能是由点燃的火柴发出的热量提供的。而对于细胞内水溶液里的分子来说，推动力则来自于与周围分子异常的极具能量的随机碰撞——一种随温度增加而更剧烈的碰撞。

活细胞内有一类特化的蛋白质——**酶**（enzyme），大大促进这种能量障碍的逾越。每种酶与一个或两个分子牢固结合，称为**底物**（substrate），并且使这些底物保持在某种形式以大大减少它们发生某一特定化学反应能。能降低反应活化能的物质称为**催化剂**（catalyst）。催化剂增加了化学反应的速率，因为催化剂能让更多比例的与周围分子的随机碰撞参与推动底物越过能量障碍。如图2-45所示，酶是已知的最有效的催化剂，能增加反应速度达10^{14}数量级。因而它们能让在常温条件下不能发生的反应快速进行。

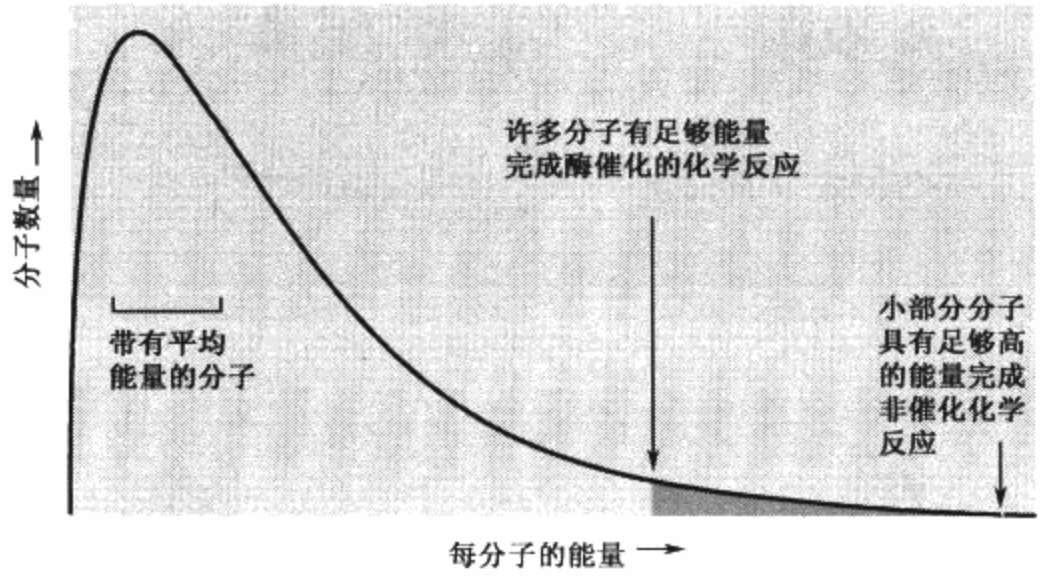


图 2-45　活化能的减少大大增加了反应进行的可能性

许多相同的底物分子具有的能量将处于一定范围内，它们在任何时刻按图中所描绘的那样分布。不断改变的能量来源于同周围分子的碰撞，这些分子使底物分子微动、振动和旋转。一个分子要进行化学反应，它的能量必须大于反应的活化能。对于大多数生物反应，如果没有酶催化作用，这是永远不会发生的。即使存在酶催作用，底物分子也必须经过一种极具能量的碰撞才能进行反应，正如这里所说的这样

酶也具有高度的选择性，每种酶通常仅仅催化一种特定的反应。换句话说，它选择性地降低了自身结合底物能可发生的许多反应中的一种的活化能。通过这种形式，酶指导细胞中每个分子沿着特定的反应途径进行（图 2-46）。

有机体存活归功于细胞产生许多种类酶的能力，这些酶每种都有独特的精确功能，而且每种酶都有独特的形状，包含一个活化位点，活化位点是一个口袋或凹槽，只有特定的底物才适合（图 2-47）。与所有其他催化剂一样，酶分子本身在参加某个反应后保持不变，因而能够反复发挥作用。在第 3 章，在我们仔细考察蛋白质分子结构之后，将进一步讨论酶是怎样起作用的。

酶是怎样找到底物的——迅速扩散极其重要

一个典型的酶每秒钟将催化大约 1000 个底物分子发生反应。这意味着每微秒内它必须能够结合到一个新的底物分子。但是无论是酶还是底物，在细胞都比较少，它们怎么能够如此迅速地找到对方呢？由于热能导致的运动在分子水平上非常迅速，所以这种快速结合成为可能，这些分子运动可以被归为三类：①分子从一个位置到另一个位置的平移运动（translational motion）；②共价相连的原子间迅速的前后移动（振动）；③旋转。所有这些运动对于相互作用的分子表面彼此接触是很重要的。

这些分子的运动率可以通过分光技术进行测量。这些意味这一个巨大的球蛋白分子以每秒 100 万次的速率绕其轴心摇摆或旋转。分子也在不停地进行平移运动。这些运动使得它们在细胞内游动时，能非常有效地开拓空间——这个过程称为**扩散**（diffusion）。通过这种方式，细胞内每个分子每时每刻都与许多其他分子相互碰撞。如果处于液体中的分子互相碰撞并弹开，那么每个单独的分子一开始沿这个方向运动，接着双向另一方向运动，它的运动路径好像随机行走出的一样（图 2-48）。在这种行走过程中，每个分子运动离起点的平均距离与所用时间的平方根呈比例。也就是说，如果一个分子平均每秒移动 1μm，那么它移动 2μm 需要 4s，移动 10μm 需要 100s 等。

细胞内的空间狭小拥挤（图 2-49），但是当把荧光染料或带有标记的分子渗入细胞内之后

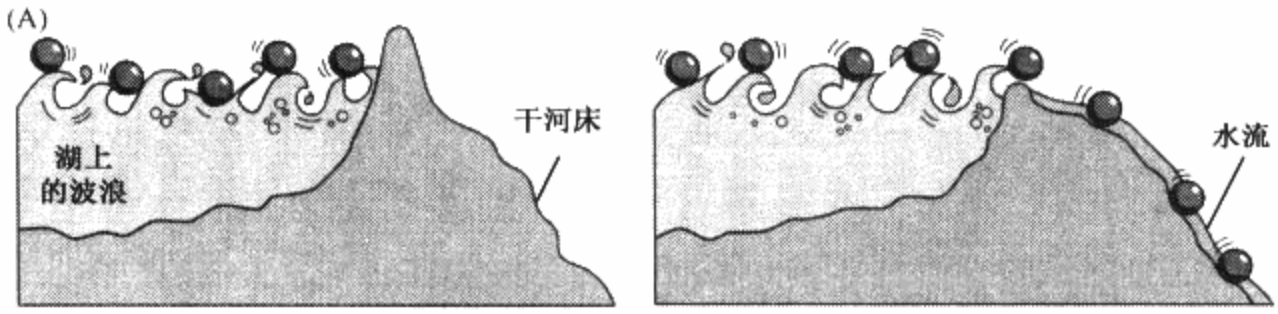


图 2-46　用浮动的球类比酶催化反应

水坝降低代表酶促反应。绿球代表可能的酶底物（复合物 X），它与波浪不断碰撞因而在能量水平上下跳动（类比底物与周围水分子之间的热撞击）。如果障碍（活化能）显著降低，就允许底物顺势而下做能量上有利的运动。（B）盒子的 4 个壁代表 4 个不同化学反应的活化能障碍，从产物能量水平较底物低的意义上来讲，这些反应都是能量上有利的反应。左边的盒子中，没有一个反应能够发生，因为即使是最大的波浪也不足以逾越任何一种能量障碍。在右边的盒子中，酶促反应降低了反应 1 的活化能。现在拥挤的波浪允许分子仅仅通过这个能量障碍，引发了反应 1。（C）用一条有许多坝的带分支的河流（黄色盒子），来说明一系列的酶促反应怎样决定细胞内每个分子确切的反应途径

发现有机小分子在细胞质溶胶中如同在水中一样迅速扩散。一个有机小分子，如扩散 10μm 远的距离需要 1/15s。因此，对于小分子在细胞内做有限距离的运动，扩散是一种很有效的

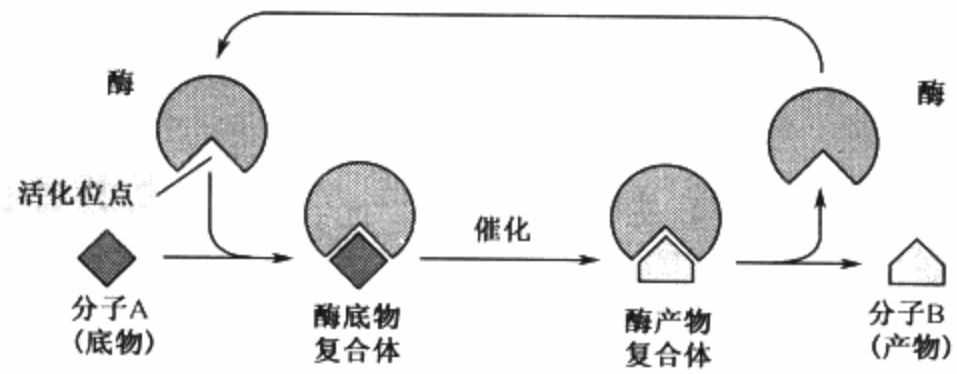


图 2-47　酶是怎样工作的

每个酶分子都有一个活化位点，结合一个或两个底物分子后，形成酶-底物复合物。A 反应发生在活化位点，生成一个酶-产物复合物。接着释放这个产物，使得酶可以结合另外的底物分子

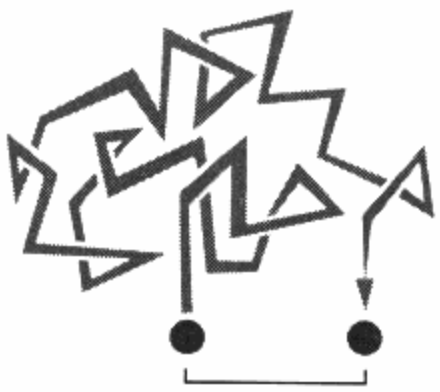


图 2-48　随机走动

溶液中的分子由于不断与其他分子发生碰撞，因而以随机的方式运动。如文中所述，这种运动使得小分子可以从细胞的一部分迅速扩散到另一部分

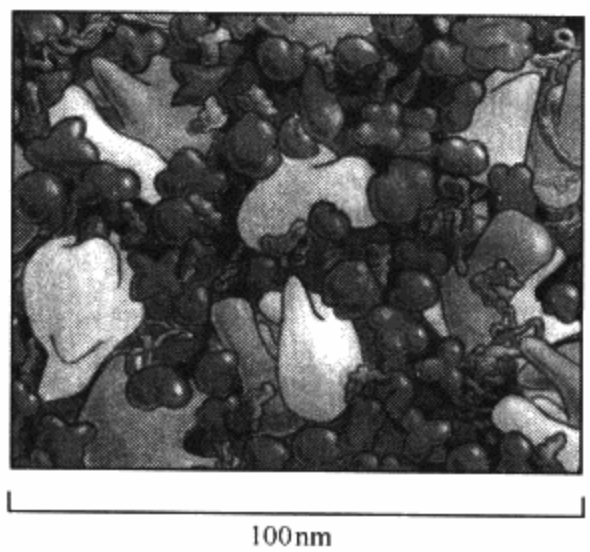


图 2-49　细胞质的结构

这幅图接近标注的比例尺，强调细胞质中的拥挤状况。图中仅仅显示了大分子：蓝色表示 RNA，绿色表示核糖体，红色表示蛋白质。酶和其他分子在细胞质中扩散相对较慢，部分原因可能是它们与许多其他分子相互作用；相比而言，小分子扩散几乎与它们在水中一样迅速（摘自 D. S. Goodsell, *Trends Biochem. Sci.* 16：203～206，1991）

方式（典型动物细胞直径为 15μm）。

细胞内酶的移动比底物慢得多，因此我们可以假定它们静止不动，那么每个酶分子遇到它的底物的概率就取决于底物分子的浓度。例如，一些含量丰富底物以 0.5mmol/L 的浓度存在。由于纯水的浓度是 55mol/L，所以细胞内每 10^5 个水分子只有大约 1 个这种底物分子。然而，每个结合这个底物的酶分子上的活性位点每秒钟与底物分子碰撞大约 500 000 次。（如果底物浓度降低 10 倍，那么碰撞的数量降低至 50 000 次/s，依此类推。）酶表面与合适的底物分子之间的碰撞，能迅速形成酶-底物复合物，并能进行反应。现在，发生共价键断裂或生成的反应就能极其迅速地进行。尽管我们惊羡于分子如此迅速地运动和反应时，观察到的酶促反应却看起来并不那么令人惊奇。

一旦酶与底物接触并在活性位点正确结合，那么它们彼此间就会形成微弱的键，这些键一直存在，直到随机热运动导致这些分子再次分离。通常酶与底物结合越牢固，解离速率就越小。然而，如果两个发生碰撞的分子不具备匹配的表面，就几乎不能形成非共价键，并且

它们的总能量相对于热运动总能量来说可以忽略不计。在这种情况下，两个分子解离就如同碰撞一样迅速。这就是防止不相匹配的分子间（如酶与非底物间）形成不正确和不需要的结合的机制。

自由能变化决定反应能否发生

现在我们必须转变话题简短介绍一些化学基础知识。细胞是化学系统，必须遵守所有的化学和物理定律。尽管酶加速反应进行，它们并不能靠自身推动能量上不利的反应发生。以水为例，酶自身并不能让水由低向高处流。但是细胞必须那样做以进行生成和分裂。它们必须利用小而简单的分子生成高度有序且富含能量的分子。我们应当看到，这必须通过酶来完成，这些酶将放能并产热的在能量上有利的反应与产生生物有序性的在能量上不利的反应偶联。

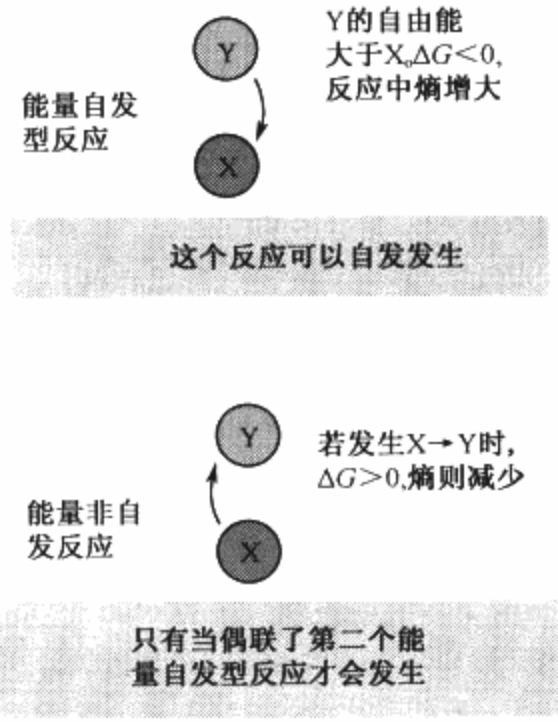


图 2-50　能量上有利和不利的反应之间的区别

在检查如何实现这种偶联之前，我们必须更加仔细地考虑“能量上有利”的含义。根据热动力学第二定律，只有当一个反应导致宇宙的无序度净增加时，它才能自发地进行(图 2-38)，宇宙无序性增加的标准可以更方便地用一个表示数量的词：系统的**自由能**（G）表示。只有当系统发生变化时，G 的值才有意义。而 G 的改变，用 ΔG 表示是非常关键的。假设把系统当作分子的集合，正如专题 2-7 所解释的那样，自由能定义的产生，使得 ΔG 直接衡量涉及这些分子的反应发生时，宇宙无序性增加的数量。根据定义，能量上有利的反应是指那些自由能减少，或者换句话说，ΔG 和宇宙自由度改变为负数(图 2-50)。

宏观上有关能量上有利的反应的一个熟悉的例子是：在弹簧从压缩状态恢复到拉长状态的“反应”过程中，以热的形式释放它们储存的能量到环境中去。微观上的一个例子是盐溶于水的反应。相反，能量上不利的反应，ΔG 为*正*。例如，两个氨基酸连接生成一个肽键的反应，自身能产生有序性。因此，这些反应只有与另一个反应偶联时，由于后者的 ΔG 如此负，以至于整个过程 ΔG 为负时才能发生（图 2-51）。

反应物浓度影响 ΔG

如上所述，反应 A↔B 只有当偶联的自由能改变 ΔG 为负时，才能沿 A→B 方向进行。就像一个绷紧的弹簧被释放时，将恢复常态并以热的形式释放储存的能量到环境中去。然而，对于一个化学反应，ΔG 不仅依赖于各个单独的分子储备的能量，而且与反应混合物中反应物的浓度有关。记住 ΔG 反映一个反应产生更加无序的，换句话说，更为可能的宇宙状态。再次提及我们用硬币做过的类比，如果一个摇动的盒子有 90 个头和 10 个尾，那么很可能硬币会从头向尾的方向翻转，但是如果盒子有 10 个头 90 个尾的话，就不大可能这样了。

正是由于相同的原因，对于一个可逆的反应 A↔B 来说，当 A 相对于 B 大大过量时，反应倾向于由 A→B 方向进行。也就是说从 A→B 转变的分子将多于从 B→A 转变的分子。因此当 A 与 B 的比值增加时，从 A→B 转变的 ΔG 变得更负（并且从 B→A 转变的 ΔG 更正）。

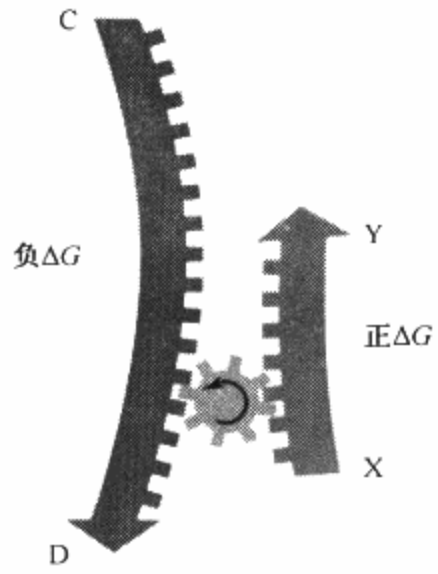


图 2-51　怎样利用偶联反应推动能量上不利的反应进行

究竟需要多少浓度的差别来弥补给定的化学键能的减少呢（伴随热能释放）？答案并非是不言自明的，而是可以从热动力学的分析中得出，这种分析可以将依赖于浓度和不依赖于浓度的自由能变化分开。因而一个给定反应的 ΔG 可以写成两个部分：第一，叫做标准自由能变化，ΔG^0，由反应物分子自身的特性决定；另一部分由反应物的浓度决定。对于 37℃条件下发生的简单反应A→B，

$$\Delta G=\Delta G^0+0.616\ln\frac{[B]}{[A]}$$

式中 ΔG 的单位是 kcal/mol；[A] 和 [B] 分别表示 A 和 B 的浓度；ln 表示自然对数；而 0.616 指 RT——气体常数 R 与绝对温度 T 的乘积。

注意当 A 和 B 的浓度相等时（ln1＝0），ΔG 等于 ΔG^0。当 B 与 A 的比值减少时（一个小于 1 的数其自然对数＜0），正如所期望的那样，ΔG 变得更负。

当浓度效应正好抵消 ΔG^0 给予反应的动力时，就达到化学**平衡**（equillibrium）了，所以没有自由能的净变化来推动反应向任何一个方向进行（图 2-52）。这里 $\Delta G=0$，所以 A 与 B 的浓度为：

$$-0.616\ln\frac{[B]}{[A]}=\Delta G^0=-1.42\ \log\frac{[B]}{[A]}$$

这表明 37℃时，当

表 2-5　标准自由能变化，ΔG^0 与平衡常数之间的关系

平衡常数 $\frac{[B]}{[A]}=K$ (L/mol)	自由能 B 减去自由能 A (kcal/mol)
10^5	−7.1
10^4	−5.7
10^3	−4.3
10^2	−2.8
10	−1.4
1	0
10^{-1}	1.4
10^{-2}	2.8
10^{-3}	4.3
10^{-4}	5.7
10^{-5}	7.1

$$\frac{[B]}{[A]}=e^{-\Delta G^0/0.616}$$

时，化学反应达到平衡状态。表 2-5 表示出 A 与 B 的平衡比例［表示为**平衡常数**（equillibrium constant），K］如何依赖于 ΔG^0 的值。

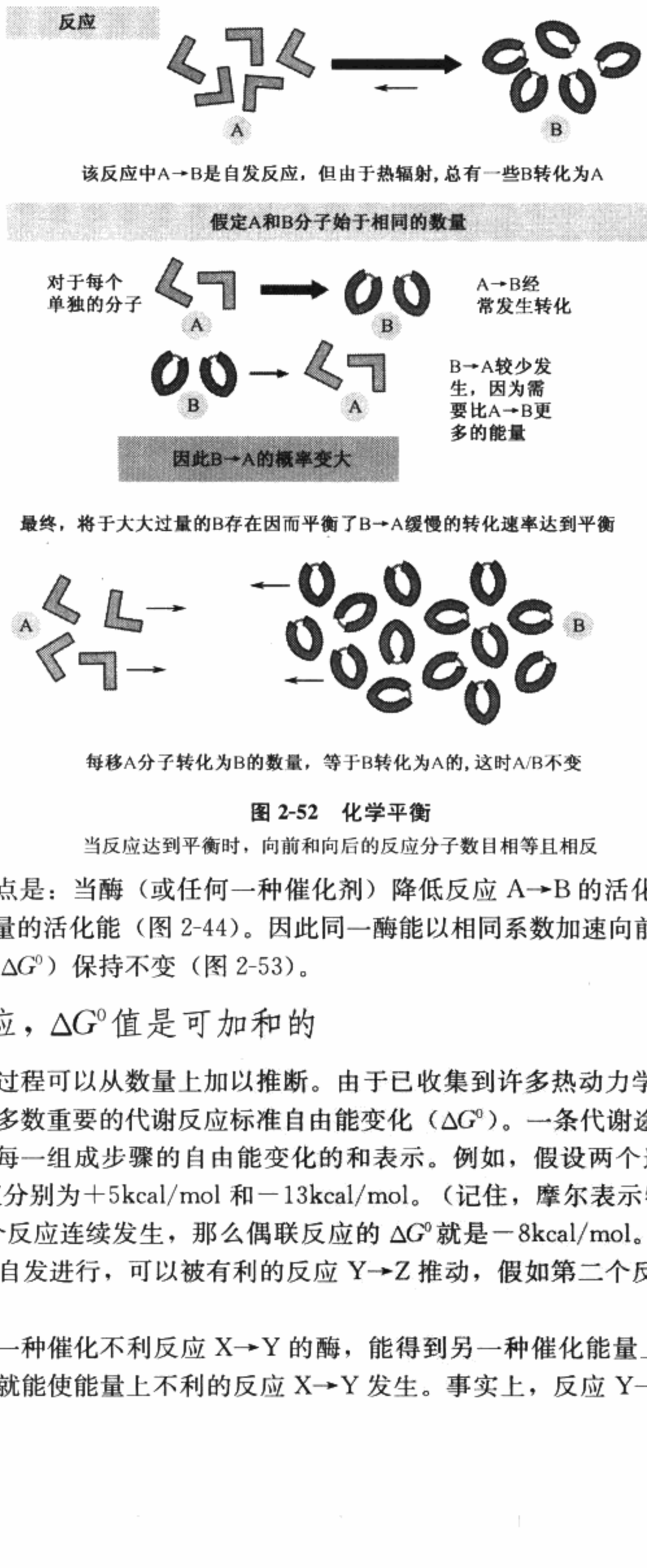


图 2-52　化学平衡

当反应达到平衡时，向前和向后的反应分子数目相等且相反

很重要的一点是：当酶（或任何一种催化剂）降低反应 A→B 的活化能时，也同时降低反应 B→A 相同量的活化能（图 2-44）。因此同一酶能以相同系数加速向前和向后的反应，反应的平衡点（和 ΔG^0）保持不变（图 2-53）。

对于连续反应，ΔG^0 值是可加和的

大部分反应过程可以从数量上加以推断。由于已收集到许多热动力学数据，所以我们能够计算细胞中大多数重要的代谢反应标准自由能变化（ΔG^0）。一条代谢途径的所有自由能变化可以简单地用每一组成步骤的自由能变化的和表示。例如，假设两个连续的反应：X→Y 和 Y→Z。ΔG^0 值分别为＋5kcal/mol 和－13kcal/mol。（记住，摩尔表示物质的 6×10^{23} 个分子。）如果这两个反应连续发生，那么偶联反应的 ΔG^0 就是－8kcal/mol。因此不利的反应 X→Y 本身并不能自发进行，可以被有利的反应 Y→Z 推动，假如第二个反应接着第一个发生的话。

因此，如果一种催化不利反应 X→Y 的酶，能得到另一种催化能量上有利的反应 Y→Z 补充的话，细胞就能使能量上不利的反应 X→Y 发生。事实上，反应 Y→Z 将接着像吸管一

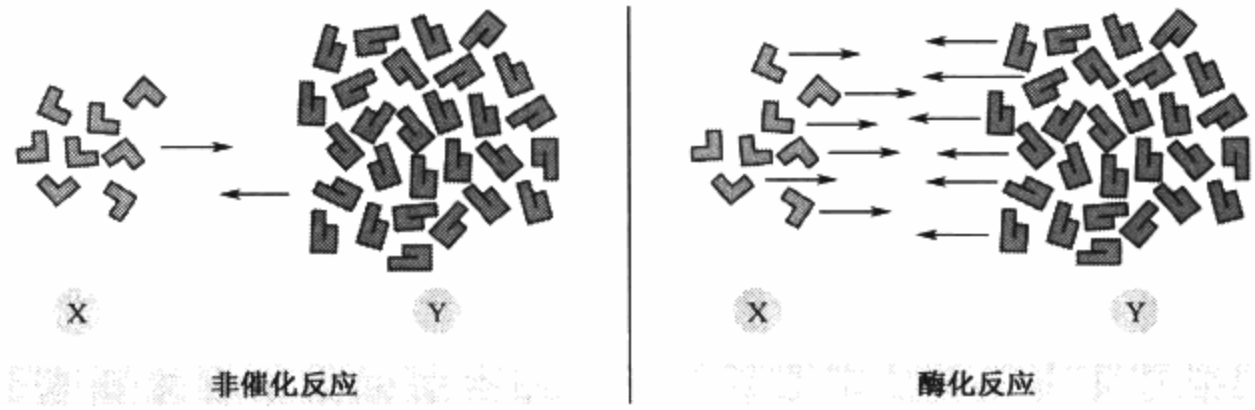


图 2-53　酶不能改变反应平衡点

如同所有的催化剂那样，酶能以相同系数改变反应向前和向后的速度。因此，在图中所示的被催化和未被催化的反应中，当分子 Y 与分子 X 的比值为 3.5∶1 时，从 X → Y 转变的分子数等于从 Y → X 转变的分子数。换句话说，两个反应在同一点达到平衡

样发挥作用，促使所有的 X 分子转变为 Y 分子，继而变为 Z 分子（图 2-54）。例如，在糖转变为 CO_2 和水的漫长途径中，许多反应如果仅从它们本身考虑，在能量上是不利的，但是由于这一系列反应的总 ΔG^0 值为很大的负值，使得该途径能迅速进行直到完成。

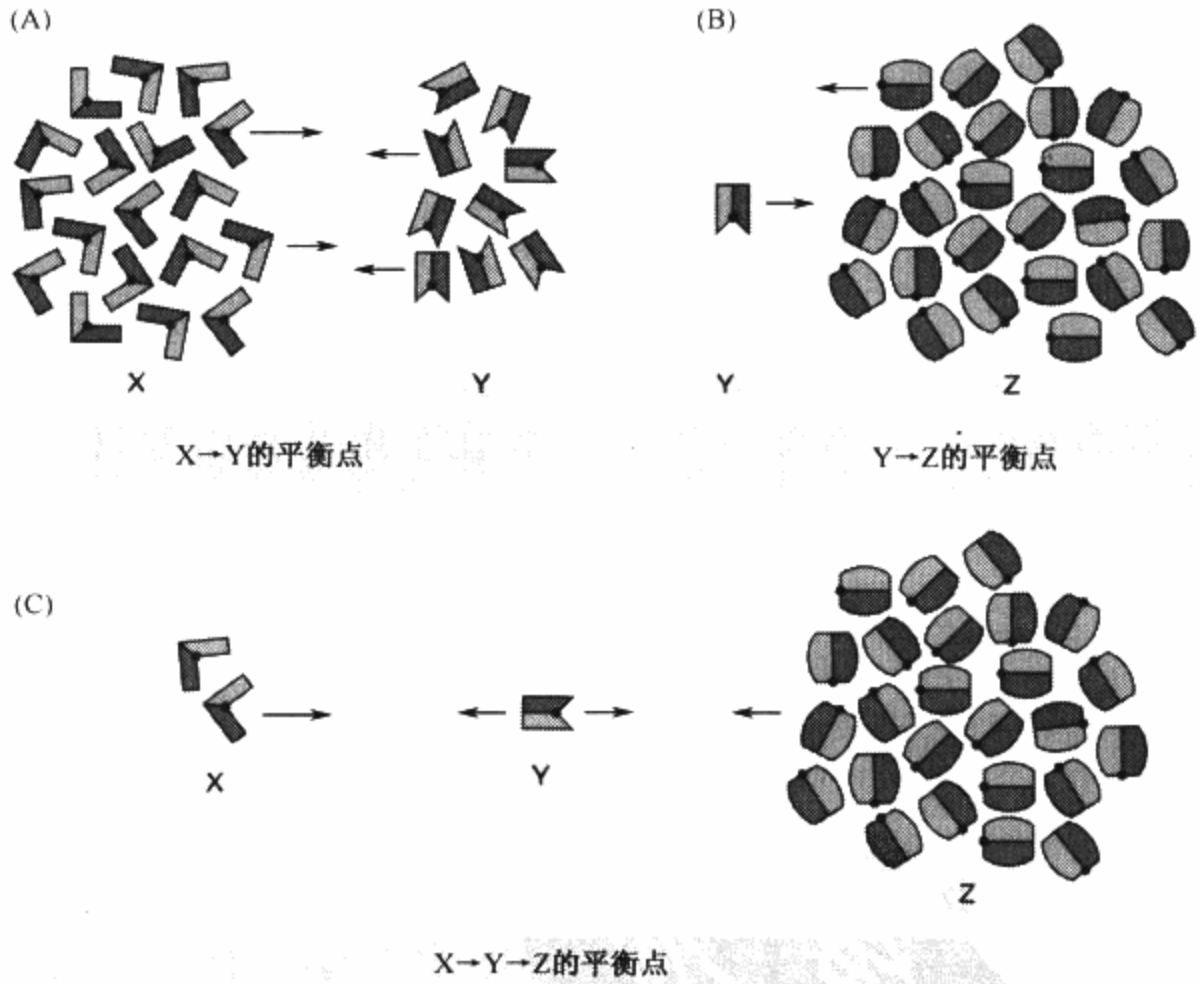


图 2-54　一个能量上不利的反应怎样被紧接着的另一个反应推动进行

（A）在平衡点，X 分子数是 Y 分子数的 2 倍，因为 X 的能量比 Y 低。（B）在平衡点，Z 分子数是 Y 分子数的 25 倍，因为 Z 的能量比 Y 低得多。（C）如果（A）和（B）中的反应偶联起来，将近所有的 X 分子将被转变成 Z 分子，如图所示

然而靠形成一个连续的反应途径并不能达到所有目的。经常发生的情况是，所需要的反应仅仅是 X→Y，并没有 Y 向其他产物的进一步转变。幸运的是，存在更多的其他通用方式来利用酶与反应偶联。这些方式是如何工作的，我们接着讨论。

活化的载体分子对于生物合成必不可少

在用于构建细胞必需的其他有机小分子和更大且复杂的分子之前，食物分子氧化所释放的能量必需暂时储存起来。在大多数情况下，这种能量以存在于一小类活化的载体分子中的化学键能的形式储存起来。这些分子含有一个或多个富含能量的共价键。这些分子在细胞内迅速扩散，因而将它们的键能从产生能量的位置运送到用于生物合成和其他细胞活动的地方（图 2-55）。

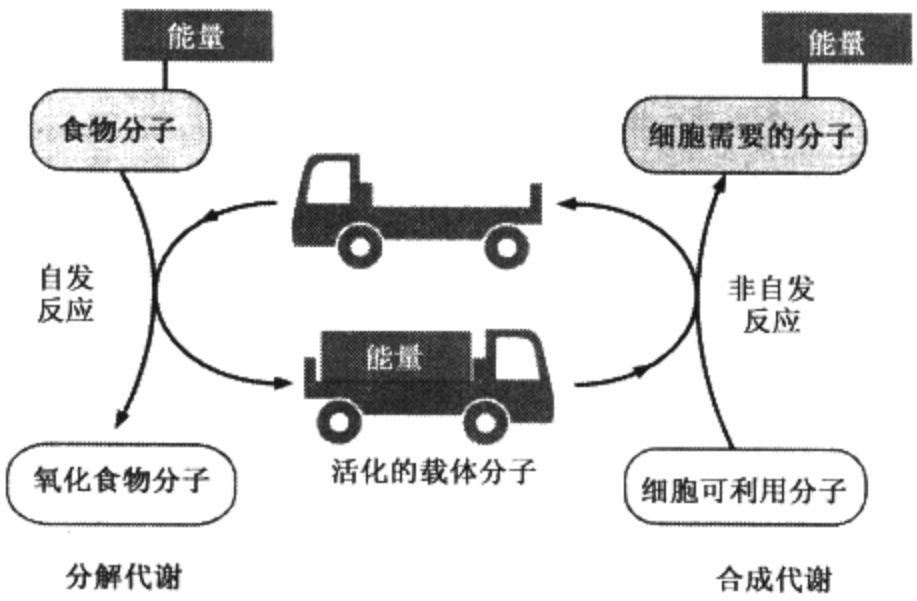


图 2-55　能量转移和活化载体在代谢中的作用

活化的载体分子作为能量的一种穿梭机制，在食物分子的降解与大小有机分子需能生物合成之间发挥媒介的作用

这些**活化载体**（activated carrier）以一种容易交换的形式储存能量，或者是易转化的化学基团，或者是高能电子。在生物合成反应中能起双重的作用，作为能量来源或者化学基团的来源。由于历史的原因，这些分子有时也称为辅酶。最重要的活化载体分子是 ATP 以及两种彼此紧密相关的分子——NADH 和 NADPH。这些我们将接着讨论。我们应当认识到细胞像利用钞票一样，利用活化的载体分子为反应“支付”能量，否则，这些反应不可能发生。

活化载体的生成与能量上有利的反应偶联

燃料分子，如葡萄糖在细胞中被氧化时，酶促反应确保了氧化释放的大部分自由能以一种化学上有用的形式被捕获，而不是以热的形式浪费掉，这一点是通过**偶联反应**（coupled reaction）实现。在反应中，能量上有利的反应推动了能量上不利的反应，后者产生一种活化载体分子或者其他的能量存储形式。偶联机制需要酶，并且对于细胞的能量交换来讲是至关重要的。偶联反应的本质通过图 2-56 中的机械的类比来加以阐明。图 2-56 中用从悬崖上滚落的石头代表能量上有利的反应。当石头撞击地面时，它的能量通常完全因摩擦产生的热的形式浪费掉（图 2-39 降落的砖块图解）。但是，通过精心的设计，这些能量部分可用于推动提升水桶的叶轮［图 2-56（B)］。由于石头现在只有推动叶轮之后才能到达地面，我们说石头下落的在能量上有利的反应与提升水桶的能量上不利的反应直接偶联起来。注意，由于部分能量用于做图 2-56（B）中的功，所以石头撞击地面的速度比图 2-56（A）中的要小，相应地就是更少的能量以热的形式浪费掉了。

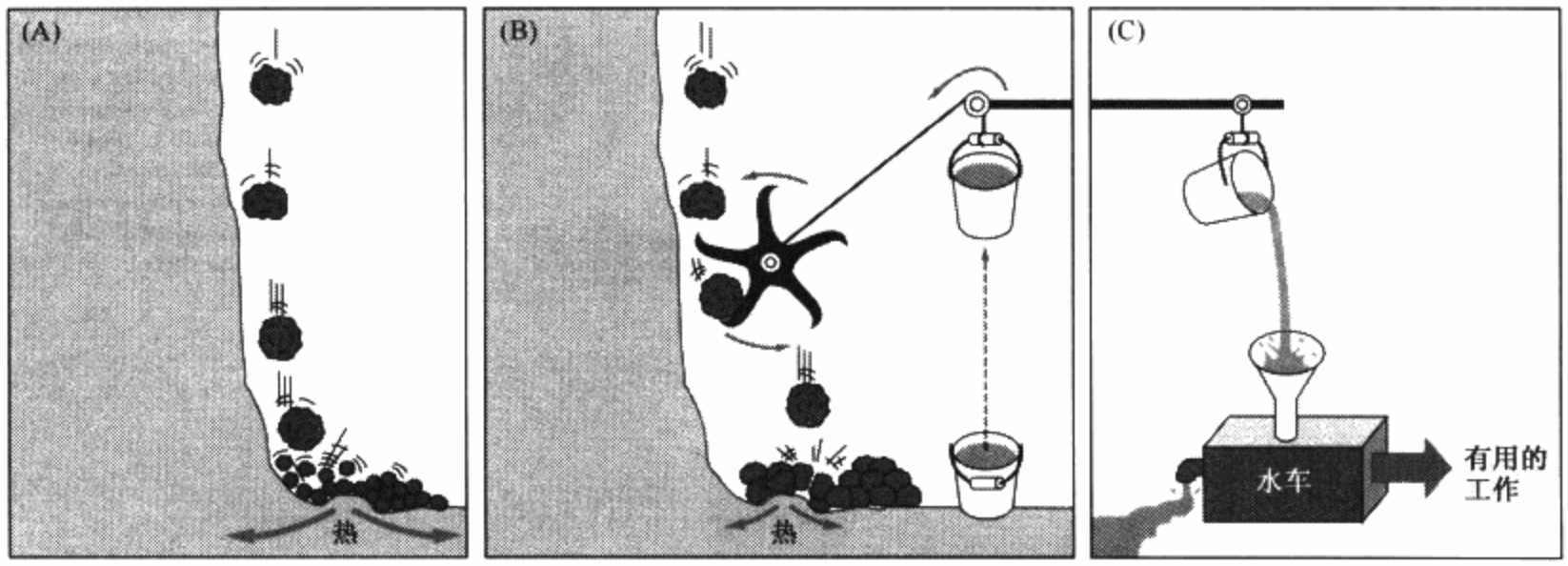


图 2-56　说明偶联反应原理的机械模型

（A）中所示的自发反应，可以用作仅仅产热的反应，葡萄糖直接氧化成 CO_2 和 H_2O 的类比。（B）中同样的反应与另一个反应偶联，这个反应可以用作合成活化载体分子的反应的类比。（B）中生成的能量可用的形式比（A）中的更多，可以用来推动各种能量上不利的反应（C）进行

恰恰类似的过程在细胞中发生了，在细胞内，酶起着叶轮的作用，通过我们本章后面将讲到的机制，这些酶使一个能量上有利的反应（如食物的氧化）与能量上不利的反应（如活化载体分子的生成）相偶联。结果氧化反应释放的热量减少，在数量上与储存于活化载体分子中的高能共价键的键能相当。活化载体分子反过来存储了许多能量，足够在细胞其他地方推动化学反应进行。

ATP 是最广泛适用的活化载体分子

细胞内最重要、功能最多的活化载体是 **ATP**（adenosine triphosphate，腺苷三磷酸）。正如图 2-56（B）中升高的水桶储存的能量可以推动各种水机器一样，ATP 在细胞中作为一种方便的和功能多样的存储形式，或者是货币或者是能量来推动各种化学反应进行。ATP 的合成是通过在 **ADP**（adenosine diphosphate，腺苷二磷酸）上加磷酸基团这种能量上不利的化学反应进行的。当需要的时候，ATP 通过能量上有利的水解作用释放能量，产生 ADP 和无机磷酸（图 2-57）。再生的 ADP 接着可用于新一轮磷酸化生成 ATP 的循环中去。

能量上有利的 ATP 水解反应与许多生成其他分子而在能量上不利的反应发生偶联。本章后面我们会涉及一些这样的反应。它们中包括了 ATP 末端磷酸转移到其他分子上，如图2-58所示的磷酸化反应。

ATP 是细胞中含量最丰富的活化载体。作为一个例子，它被用于将物质泵入或者泵出细胞的反应中提供能量，它也为分子泵提供能量，这些分子能使肌肉细胞收缩，也能使神经细胞将信息物质沿着长长的轴从一端传向另一端（讨论见第 16 章）。

储存于 ATP 中的能量通常用于两个分子的接合

前面我们已讨论过一个能量上有利的反应能够与一种能量上不利的反应 X→Y 相偶联，从而使其能够进行反应。在这种方式中，一个酶催化能量上有利的反应 Y→Z，这个过程推动

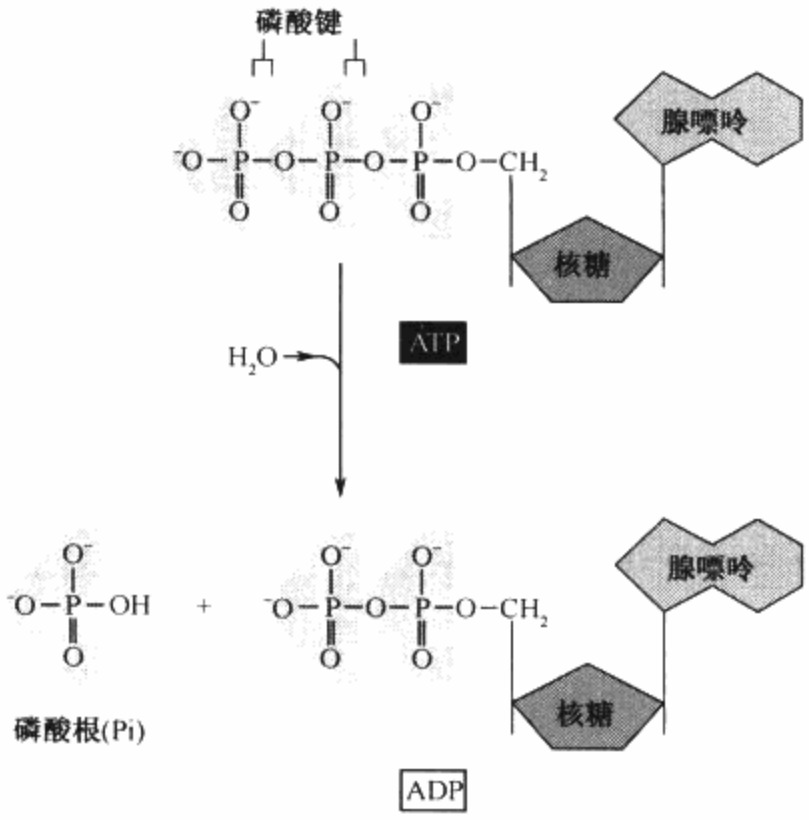


图 2-57　ATP 水解成 ADP 和无机磷酸

ATP 中最靠外的两个磷酸与其余部分通过高能磷酸键相连，因而容易转移。如文中所述，ATP 可以加水生成 ADP 和无机磷酸（Pi）。根据细胞内条件，这个末端磷酸的水解生成 11～13kcal/mol 的有用能量。许多因素导致产生巨大的负的 ΔG 值，末端磷酸的释放消除了相邻负电荷间不利的排斥；此外，释放的无机磷酸通过共振和水形成有利的氢键得以稳定

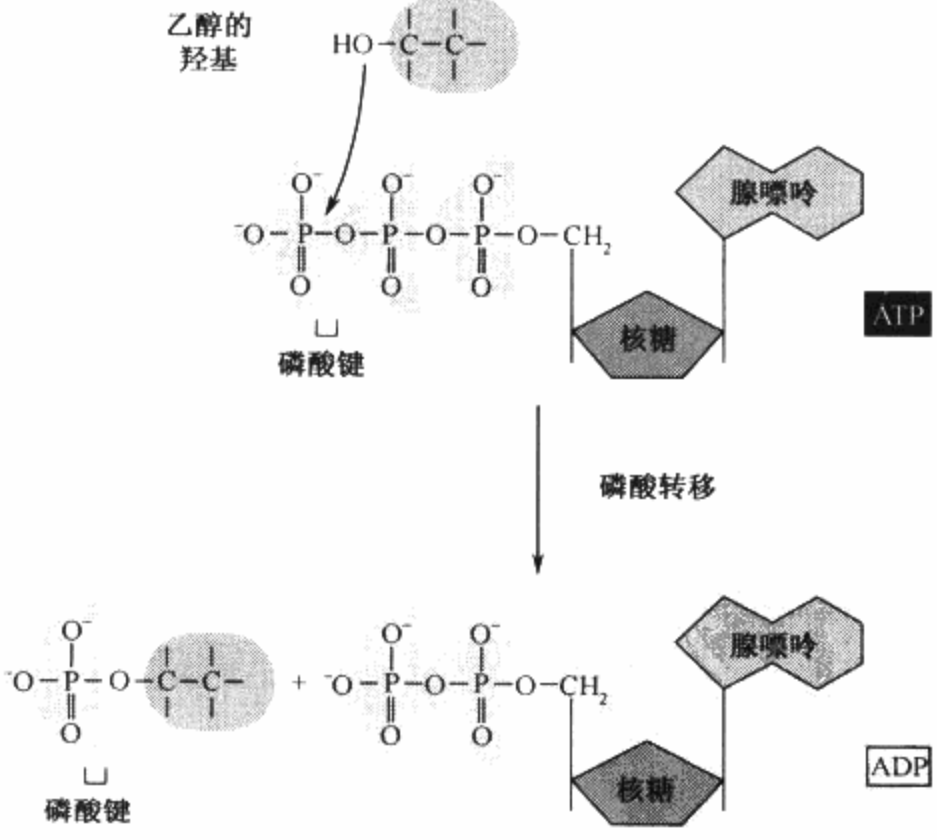


图 2-58　磷酸转移反应实例

由于富含能量的磷酸酐键转移到磷酯键上，所以这个反应在能量上有利，具有巨大负的 ΔG 值。这样的反应发生在磷酯合成以及合成糖的起始步骤

所有 X 转化为 Y（图 2-54），但是当所需的产物是 Y 而不是 Z 的时候，这种机制就没作用了。

生物合成需要经常发生的反应类型是：在能量上不利的缩合反应中，两个分子 A 和 B 连接起来生成 AB：

$$A—H+B—OH \rightarrow A—B+H_2O$$

存在一种间接途径，使 A—H 和 B—OH 生成 A—B，在这种途径中由于与 ATP 水解反应偶联，使得反应得以进行。这里 ATP 水解释放的能量首先用于 B—OH 转变为一种高能中间复合物，接着这个中间复合物直接与 A—H 反应生成 A—B。这种最为简单可行的机制涉及磷酸从 ATP 转移至 B—OH 生成 $B—OPO_3$。在这种情况下，反应途径仅仅包括两步：

1. $B—OH+ATP \rightarrow B—O—PO_3+ADP$

2. $A—H+B—O—PO_3 \rightarrow A—B+Pi$

净结果：$B—OH+ATP+A—H \rightarrow A—B+ADP+Pi$

在酶催反应途径中，本身在能量上不利的缩合反应由于直接与 ATP 水解反应偶联，因而迫使其能够进行［图 2-59（A）］。

这种类型的生物合成反应可用于合成谷氨酸，如图 2-59（B）所示。后面我们将会看到非常类似（但更复杂）的机制用于生成细胞中几乎所有的大分子。

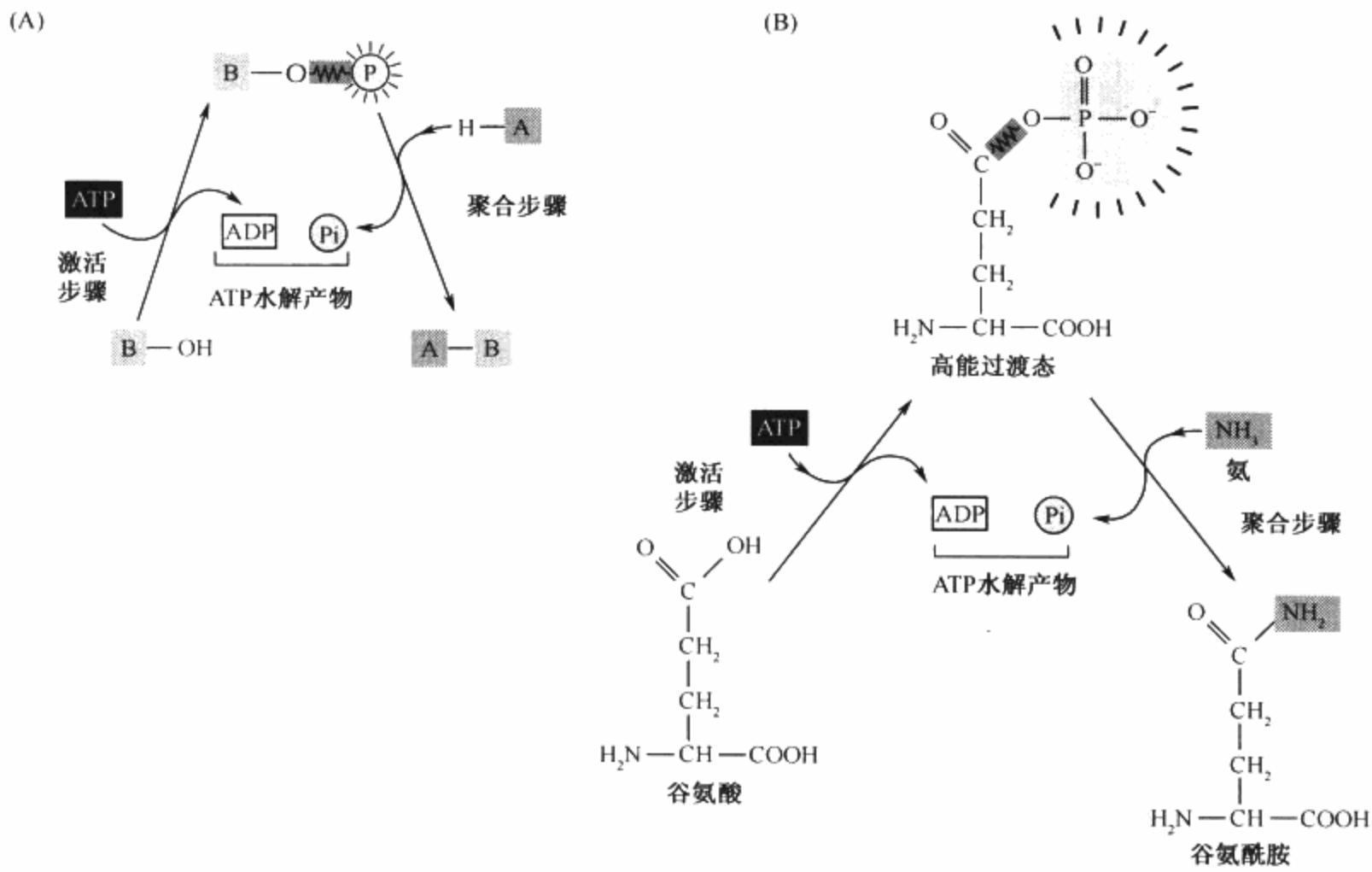


图 2-59　由 ATP 水解反应推动的一个能量上不利的生物合成反应的实例

（A）文中描述的缩合反应中 A—B 的形成的概图。（B）谷氨酰胺的生物合成。谷氨酸首先转变成高能磷酸化的中间产物（对应于文中描述的复合物 $B—O—PO_3$），接着这个中间物与氨（对应于 A—H）反应生成谷氨酰胺。本例中的两步反应在同一酶谷氨酸合成酶的表面发生。注：为了清楚起见，氨基酸用未带电荷的形式表示

NADH 和 NADPH 是重要的电子载体

其他重要的活载体分子也参与了氧化-还原反应，并且通常是细胞中的偶联反应，这些活化载体分工特化后专门运载高能中子和氢原子。这些电子载体中最重要的就是 NAD^+（烟酰胺腺

嘌呤二核苷酸）和密切相关的 $NADP^+$（烟酰胺腺嘌呤二核苷酸磷酸），我们在后面探讨它们参与的一些反应。NAD^+ 和 $NADP^+$ 分别接受 2 个高能电子和 1 个氢原子的能量——分别被转化为 NADH（还原型烟酰胺腺嘌呤二核苷酸）和 NADPH（还原型烟酰胺腺嘌呤二核苷酸磷酸）。因此这些分子也可以被认为是氢离子载体（H^+ 加上 2 个电子，或本身是 H^-）。

与 ATP 一样，NADPH 作为活化载体参与了许多重要的生物合成反应，而不存在 NADPH 时，这些反应在能量上讲是不利的。NADH 按照图 2-60（A）中所示的通用途径产生，在一系列特殊的放能催化反应中，1 个氢原子和 2 个电子从底物分子被释放出来，然后加到 $NADP^+$ 的烟酰胺环上形成 NADPH，这是一个典型的氧化-还原反应，底物被氧化，$NADP^+$ 被还原，$NADP^+$ 和 NADPH 的结构如图 2-60（B）所示。

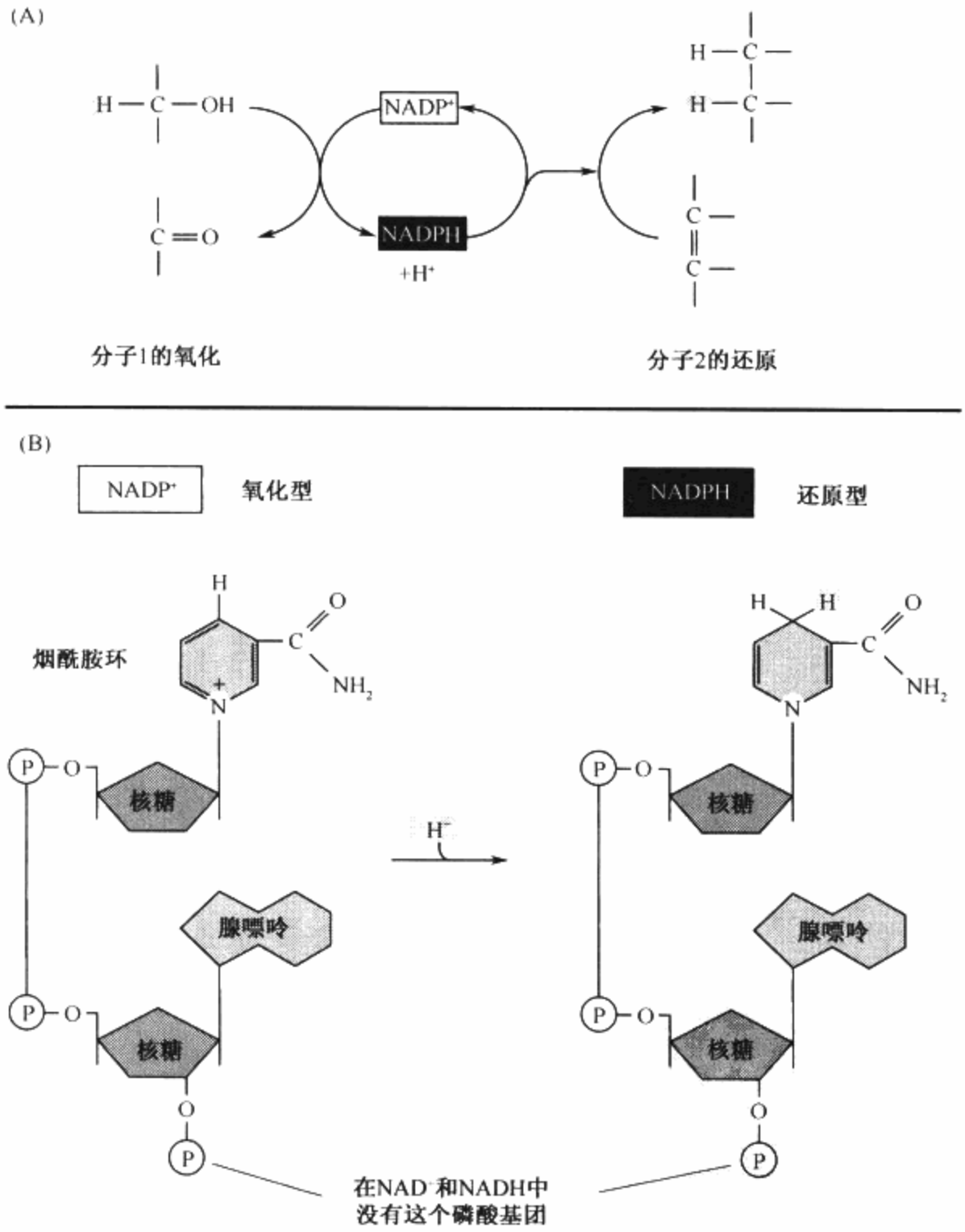


图 2-60　重要的电子载体 NADPH

（A）NADPH 通常在图中左边所示的反应类型中产生。在这类反应中，从底物移去 2 个氢原子。氧化型的载体分子 $NADP^+$ 接受 1 个氢原子和 2 个电子（1 个氢离子）而另外 1 个氢原子以质子（H^+）的形式释放到溶液中。因为 NADPH 以高能连接固定氢负离子，所以 NADPH 中的氢负离子很容易转移到其他的分子，如图中右面所示。（B）$NADP^+$ 和 NADPH 的结构。$NADP^+$ 结构中的烟酰胺环部分接受 2 个电子和 1 个质子（即氢负离子，H^-），形成 NADPH。除了两者缺少磷酸基团以外，NAD^+ 和 NADH 在分子结构上分别与 $NADP^+$ 和 NADPH 相同

NADPH 携带的氢原子很容易在随后的氧化还原反应中释放出来，因为没有它的存在，酰胺环能使电子处于更稳定的状态。这个随后发生的生成 $NADP^+$ 的反应中，NADPH 被氧化，而底物则被还原，NADPH 能有效地为其他分子提供氢离子，这与 ATP 很容易转移磷酸的道理是一样的：两种情况下的转移都伴随着一个很负的自由能交换。图 2-61 列出一个在生物合成反应中利用 NADPH 的实例，与 NADH 相比，NADPH 上的一个磷酸基团的差别并不影响传递电子的动能，但是对于二者所发挥的不同作用来说，这个磷酸基团却是至关重要的。NADPH 上额外的磷酸基团远离参与电子转移反应的区域［图 2-60（B)］，因而对于转移反应无关紧要。然而，它却使 NADPH 分子具有与 NADH 稍微不同的形状，所以 NADPH 和 NADH 作为底物能与不同类的酶结合，因此这两种载体可用于不同类的分子间电子转移反应（或氢离子)。

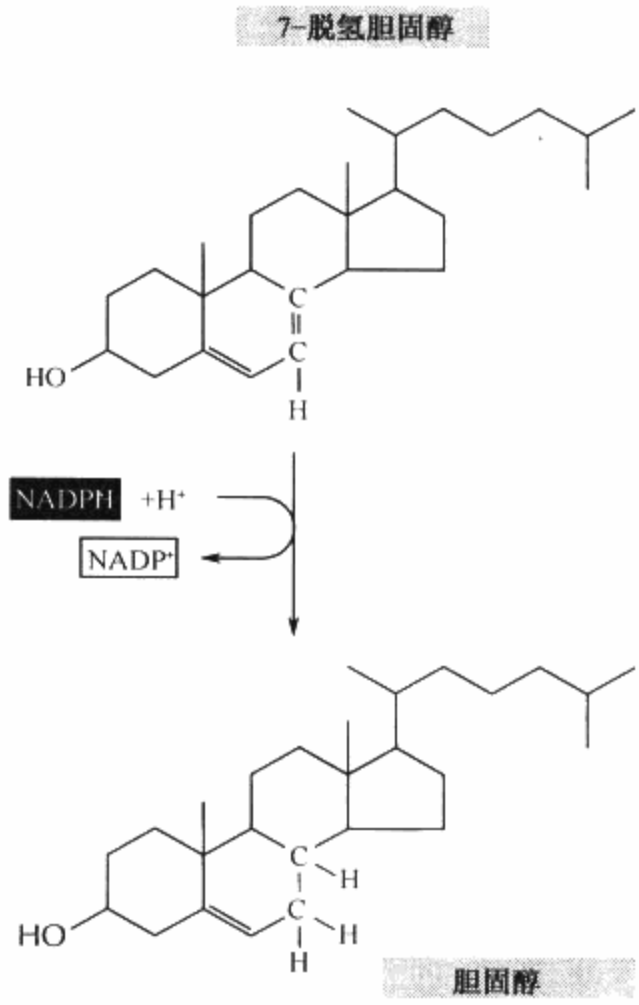


图 2-61　胆固醇生物合成途径中最后的阶段

与许多其他的生物合成反应一样，C=C 键的还原可以通过载体分子 NADPH 上转移氢负离子以及从溶液中转移一个质子（H^+）来实现

为什么会有这样的分工呢？这是因为需要对两类电子转移反应进行独立的调节，NADPH 主要与催化合成反应的酶一起作用，提供合成富含能量的生物分子时所需的高能电子。相反，NADH 作为媒介物，在食物分子氧化生成 ATP 的分解反应系统中发挥特殊的作用，这些我们将会在后面讨论。由 NAD^+ 合成 NADH 以及由 $NADP^+$ 生成 NADPH 分别在不同的途径进行，并且各自独立得到调节，所以细胞可以对立的调节电子供应，以达到两种不同的目的。在细胞内，NAD^+ 与 NADH 的比值保持较高水平，而 $NADP^+$ 与 NADPH 的比值则维持在较低的水平，这样既可以提供大量的 NAD^+ 作为氧化剂，又可以提供大量的 NADPH 作为还原剂，正如它们在分解代谢和合成代谢中各自发挥独特作用所要求的那样。

细胞内还有许多其他的活化载体分子

其他的活化载体同样接受并转运一个容易转移的高能化学基团（表 2-6)。例如，辅酶 A 携带一个易于转移的乙酰基团，这种活化的形式称为**乙酰 CoA**（acetyl CoA，乙酰辅酶 A)。图 2-62 显示的是乙酰 CoA 的结构，它主要用于在合成大分子中增加两个碳单位。

表 2-6　新陈代谢中常用的活化载体分子

活化载体	高能连接中的基团
ATP	磷酸基
NADH，NADPH，$FADH_2$	电子和氢
乙酰 CoA	乙酰基团
羧酸生物素	羧基
腺苷甲硫氨酸	甲基
二硫酸尿苷葡萄糖	葡萄糖基

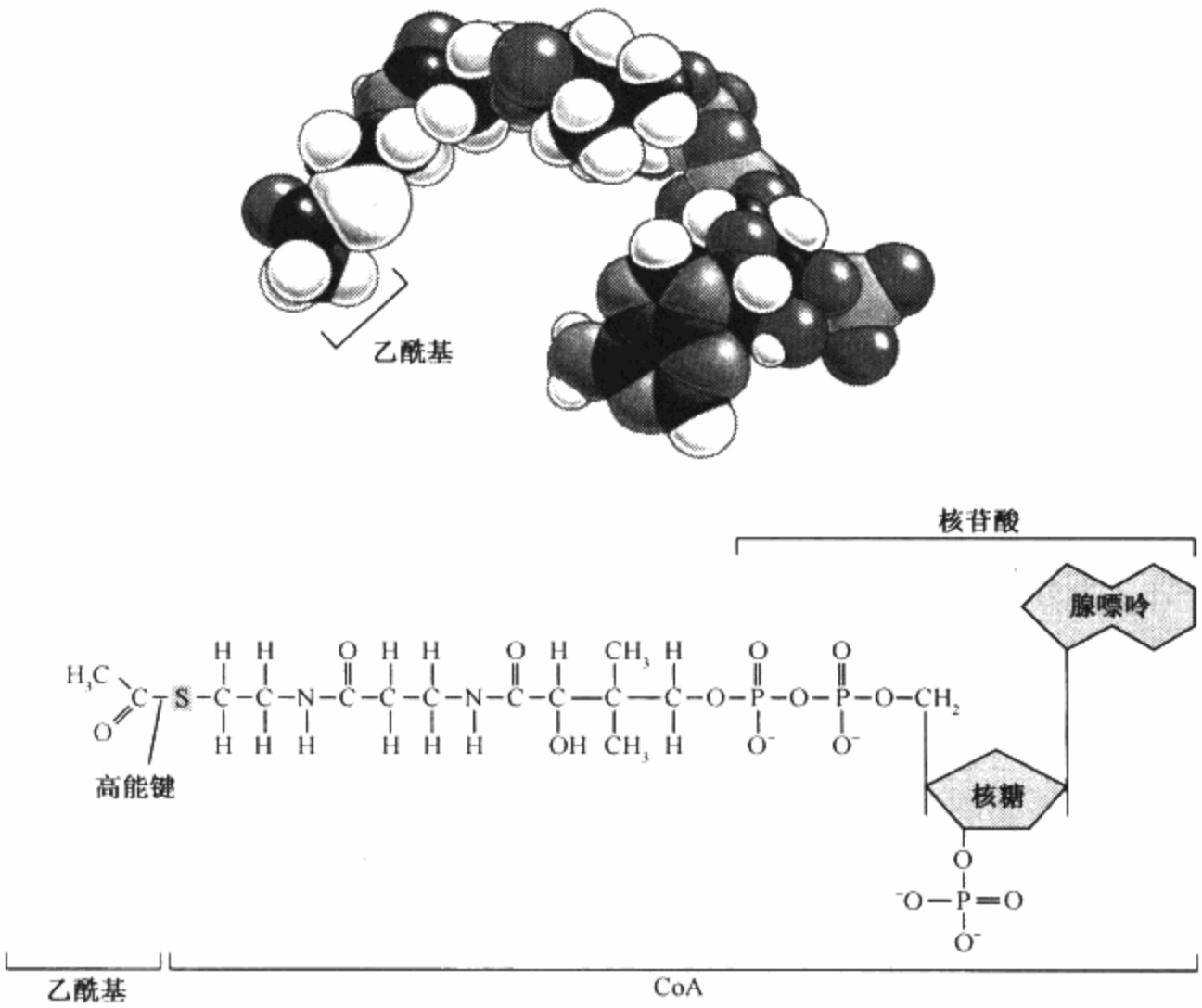


图 2-62　重要的活化载体分子乙酰 CoA 的结构

结构图上面是空间填充模型。硫原子（黄色）以硫酯键结合到酸上。因为硫酯键是一个高能键，水解时能释放大量的能量，所以乙酸分子很容易转移到其他分子上

在表 2-6 中的乙酰 CoA 以及其他载体分子中，易转移的基团只占整个分子很小的一部分，其余较大的有机部分像一个握起来方向的“木柄”一样发挥作用，以帮助特定的酶识别载体分子。正如与乙酰 CoA 中的情况一样，这个把柄部分通常含有一个核苷酸，这是一个有趣的现象，可能是早期进化的“遗迹”。目前人们认为，早期生命形式的主要催化剂——在 DNA 和蛋白质出现以前是 RNA 分子（或者与它们很相似的分子），这些在第 6 章描述。这很容易让人推测，许多我们现在发现的载体分子起源于早期的 RNA 世界，在这里它们的核苷酸部分用于结合到 RNA 酶上。

图 2-58 和图 2-61 分别列出了由 ATP（转移磷酸基团）和 NADPH（转移电子和氢）两种载体分子催化的转移反应类型。为了生物合成反应，其他活化载体分子的反应需要甲基、羧基或者葡萄糖基团的转移。反应所需的活化载体分子通常由与 ATP 水解偶联的反应生成，如图 2-63 所示。因此，使活化载体分子所携带的基团能够用于生物合成的能量最终来源于 ATP 的分解代谢反应。类似的过程也在细胞内大分子的合成中发生，如核酸、蛋白质、多糖，我们接着讨论。

生物聚合物的合成需要输入能量

正如前面讨论的，细胞的大分子占其干物质（没有水的物质）的大部分（图 2-29）。这些分子由亚基（或者单体）通过缩合反应连接起来产生，在这种反应中一个水分子的组分（OH 加上 H）从两个反应物中释放。因此，逆反应（所有三类聚合物的分解）通过酶促加水

（水解）进行，这种水解反应在能量上有利，而生物合成反应需要输入能量并且更加复杂（图 2-64）。

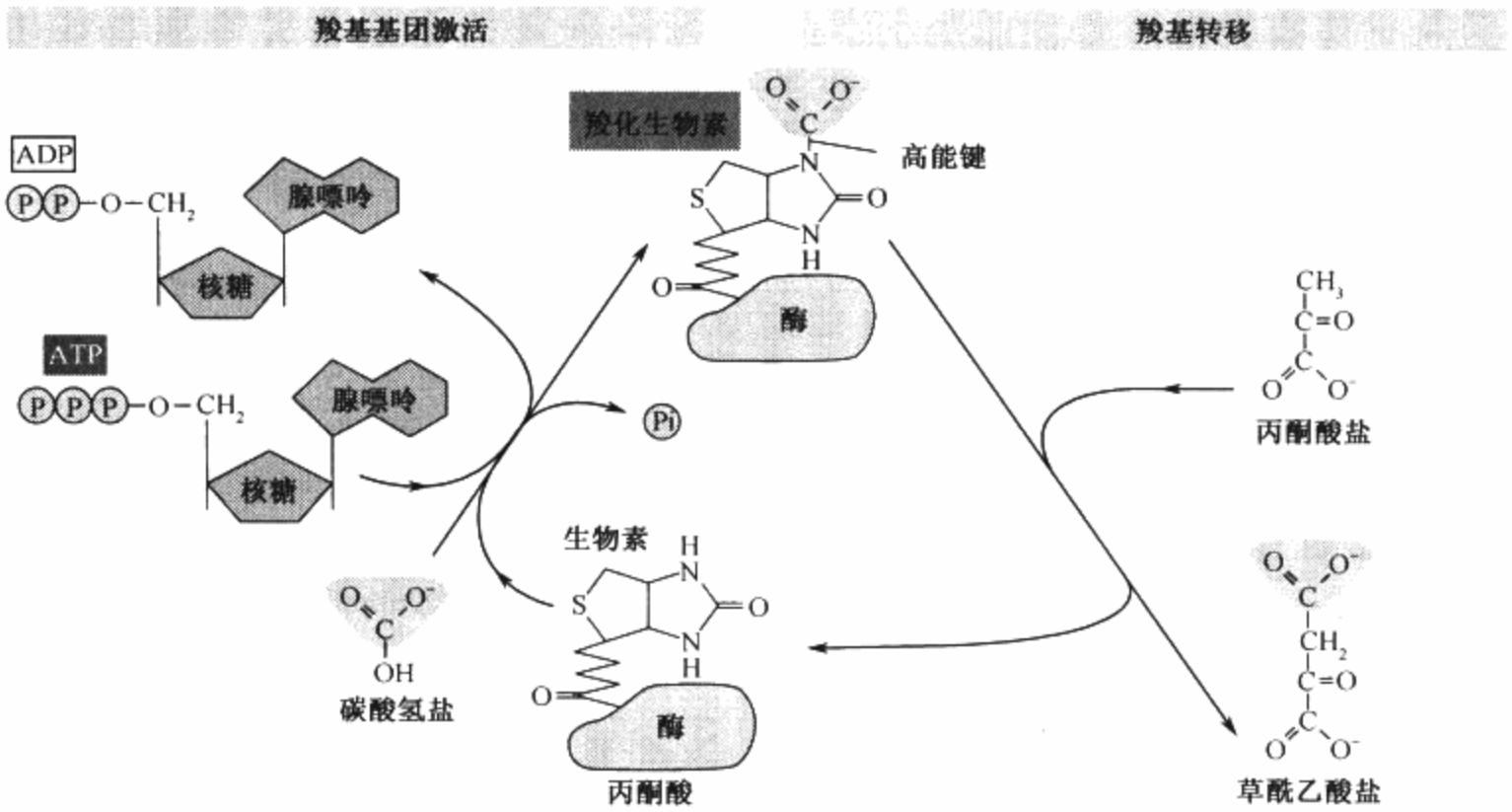


图 2-63　利用活化载体分子进行的羧基转移反应

丙酮酸羧化酶利用羧化的生物素转移羧基生成草酰乙酸，草酰乙酸是柠檬酸循环必需成分之一。羧基转移反应的受体分子是丙酮酸，其他的酶也利用生物素转移羧基到其他的受体分子上。注意：生成羧基化的生物素所需的能量来源于 ATP，这是许多活化载体的普遍特征

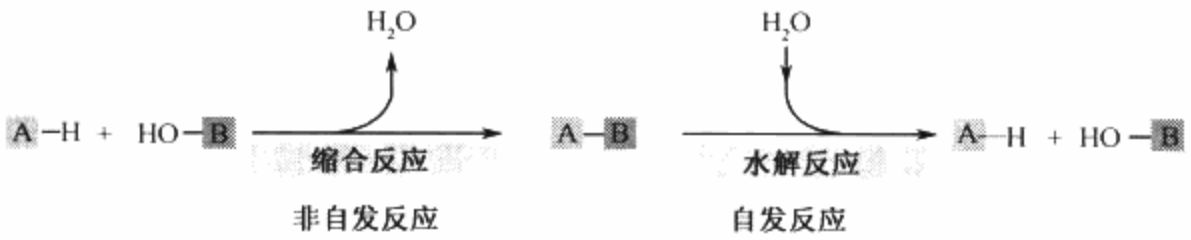


图 2-64　缩合反应与水解反应是相反方向进行的反应

细胞内的大分子都是多聚物，它们通过缩合反应由亚基（成单体）合成，并经过水解反应降解。缩合反应都是能量上不利的反应

核酸（DNA 和 RNA）、蛋白质和多糖都是通过在不断延伸的引物链末端加上亚基（也称为单体）合成的多聚物。图 2-65 概括了这三类大分子的合成反应，每一种反应的缩合步骤都依赖于核苷三磷酸水解释放的能量。但是除了核酸以外，在最终的产物分子中并无磷酸基团存在，ATP 水解释放能量与多聚体合成反应偶联是怎样进行的呢？

对于每一种类型的大分子而言，都存在一种与前面讨论的谷氨酸合成类似的酶催反应途径（图 2-59）。原理都是一样的，也就是在缩合反应中将要被移走的 OH 基团，通过参与另一个分子高能键的形成过程而首先被活化。然而，将 ATP 水解与蛋白质和多糖合成偶联起来的确切机制比起谷氨酸合成过程更复杂，因为需要一系列高能中间产物来生成最后在缩合反应步骤中被断裂的高能键（讨论见第 6 章蛋白质合成）。

每种活化载体在推动生物合成反应中发挥的作用也很有限。ATP 分解为 ADP 和无机酸（Pi）反应的 ΔG 依赖于所有反应物的浓度。但在通常情况下，细胞内自由能变化为

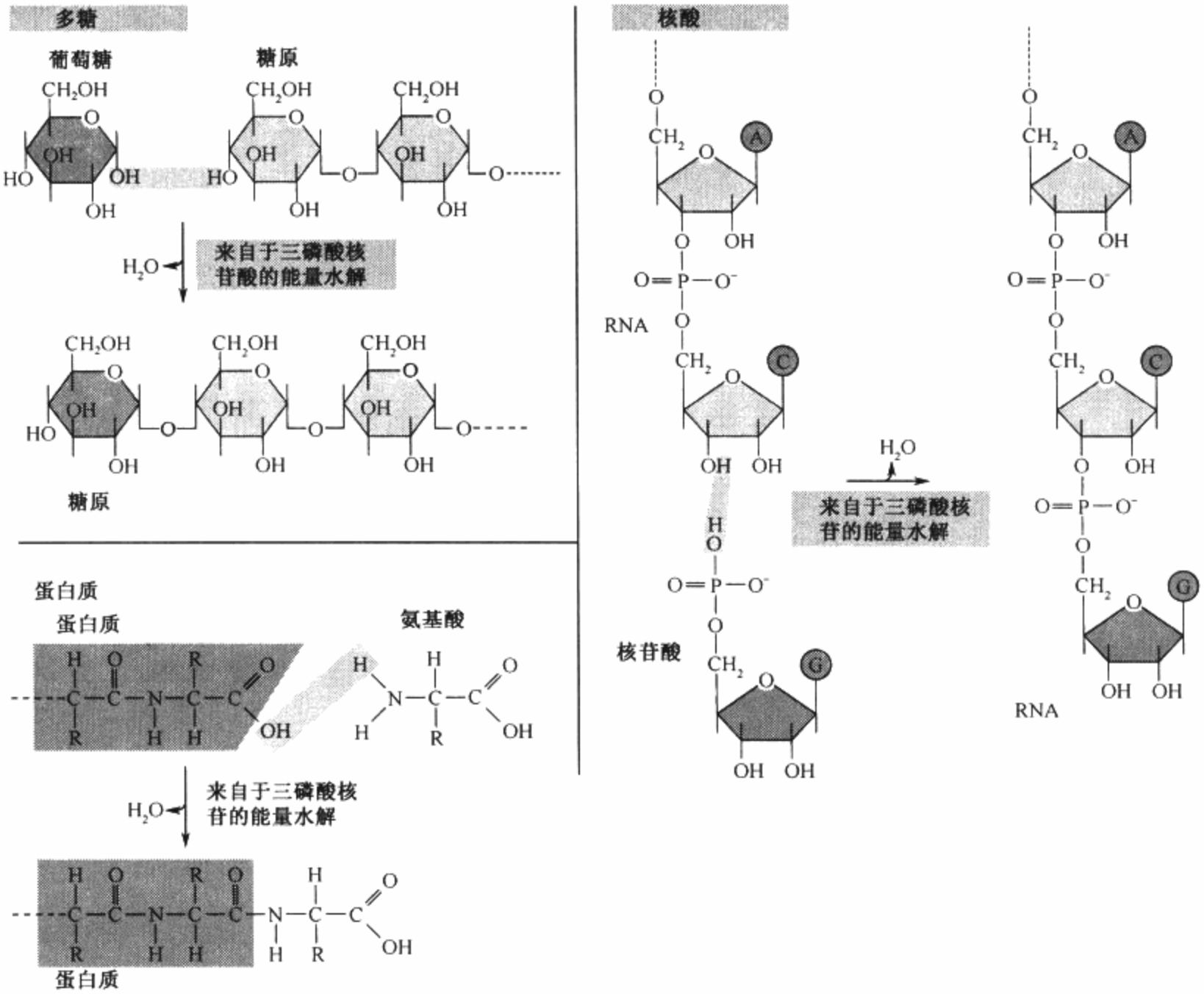


图 2-65　多糖、蛋白质以及核酸的合成

每一种生物多聚体的合成都涉及缩合反应中水分子的丢失。图中没有显示出在加成之前，每一个单体分子活化所需的高能核苷三磷酸的消耗。相比之下，逆反应——所有三类多聚体的分解，通过简单的加水（水解）进行

−11～−13kcal/mol。理论上，如果存在合适的反应途径，这个水解反应可以用来推动到一个 ΔG 大约为+10kcal/mol 的不利反应。但是，对于某些生物合成反应，即使−13kcal/mol 也可能不够。在这些情况下，ATP 水解途径可以发生改变，这样起初 ATP 水解生成 AMP 和焦磷酸（PP_i），然后焦磷酸水解（图 2-66），整个过程总共可以获得大约−26kcal/mol的自由能变化。通过这种方式推动发生的一个重要生物合成反应是核酸（多聚核苷酸）的合成，如图 2-67 所示。

有意思的是，生成大分子的聚合反应可以向两个不同的方向中的一种进行，从而使得单体向头部聚合或者沿尾部聚合。头部聚合反应中，缩合反应需要的活化键由生长的聚合物末端携带，因而每次单体加入的时候，它都必须再生。在这种情况下，每个单体携带的活化键将用于把下一个单体加入到序列中去。相反，在尾部聚合反应中，每个单体携带的活化键立即用于自身的加成反应（图 2-68）。

我们在后面的章节中将看到这两种聚合类型都会用到。例如，核酸和一些简单多糖的合成通过尾部聚合进行，而蛋白质合成则通过头部聚合进行。

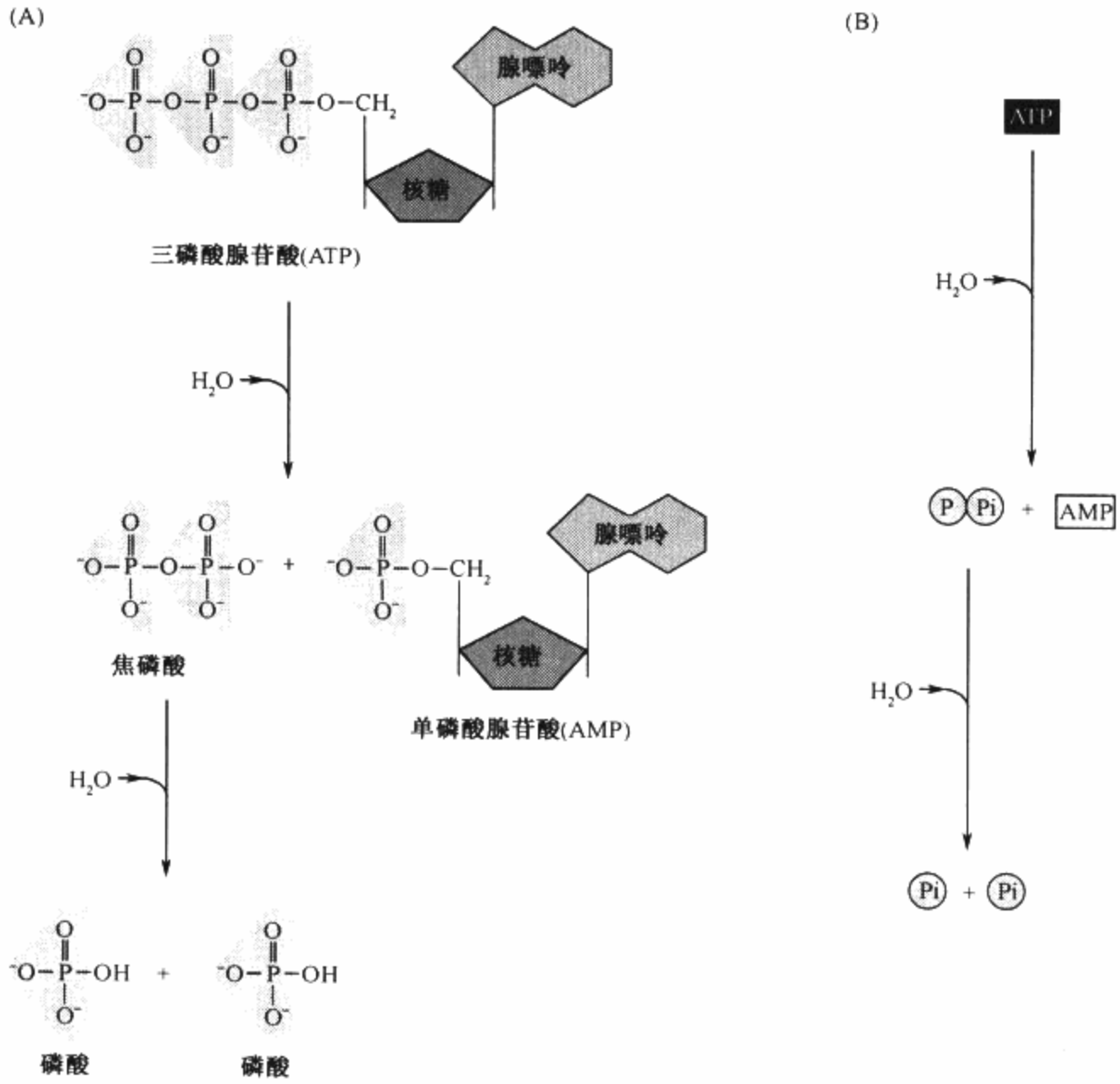


图 2-66　ATP 水解的另一条途径，该途径中首先生成焦磷酸，然后水解焦磷酸

这条途径释放的自由能大约是前面图 2-57 所示的反应的 2 倍。（A）在两个连续进行的水解反应中，来自参与反应的水分子中的氧原子保留在产物中，如图所示，而氢原子解离出来形成自由氢离子（H^+，图中未显示）。（B）整个反应过程的概图

小结

活细胞是高度有序的，为了存活和生长，它们需要在自身内部建立有序性。这之所以在热力学上成为可能，仅仅是因为不断吸收能量，而能量的一部分作为热能从细胞释放到环境中去。能量最终来源于太阳的电磁辐射，这些辐射推动在光合生物如绿色植物中生成有机分子。动物依靠采食这些有机分子，并经过一系列酶催反应氧化它们而获得能量，这些反应则与 ATP 形成相偶联，ATP 是所有细胞内的一种通用能量货币。

为了使细胞内能够不断产生有序性，能量上有利的 ATP 水解反应与能量上不利的反应偶联。在大分子生物合成过程中，通过转移磷酸基团形成活化的磷酸化中间产物来实现这种偶联。由于能量上不利的反应现在变得有利了，因此认为 ATP 水解反应推动了反应的进行。多聚体分子如蛋白质、核酸和多糖由大大小小的活化前体分子通过重复进行的缩合反应拼装而成。这些缩合反应就是通过以上的方式得到推动的。其他的一些活化分子，或者称为活化载体或辅酶，在生物合成反应中转移其他的化学基团。例如，NADPH 转移 1 个质子和 2 个电子（1 个氢离子），而乙酰 CoA 转移 1 个乙酰基团。

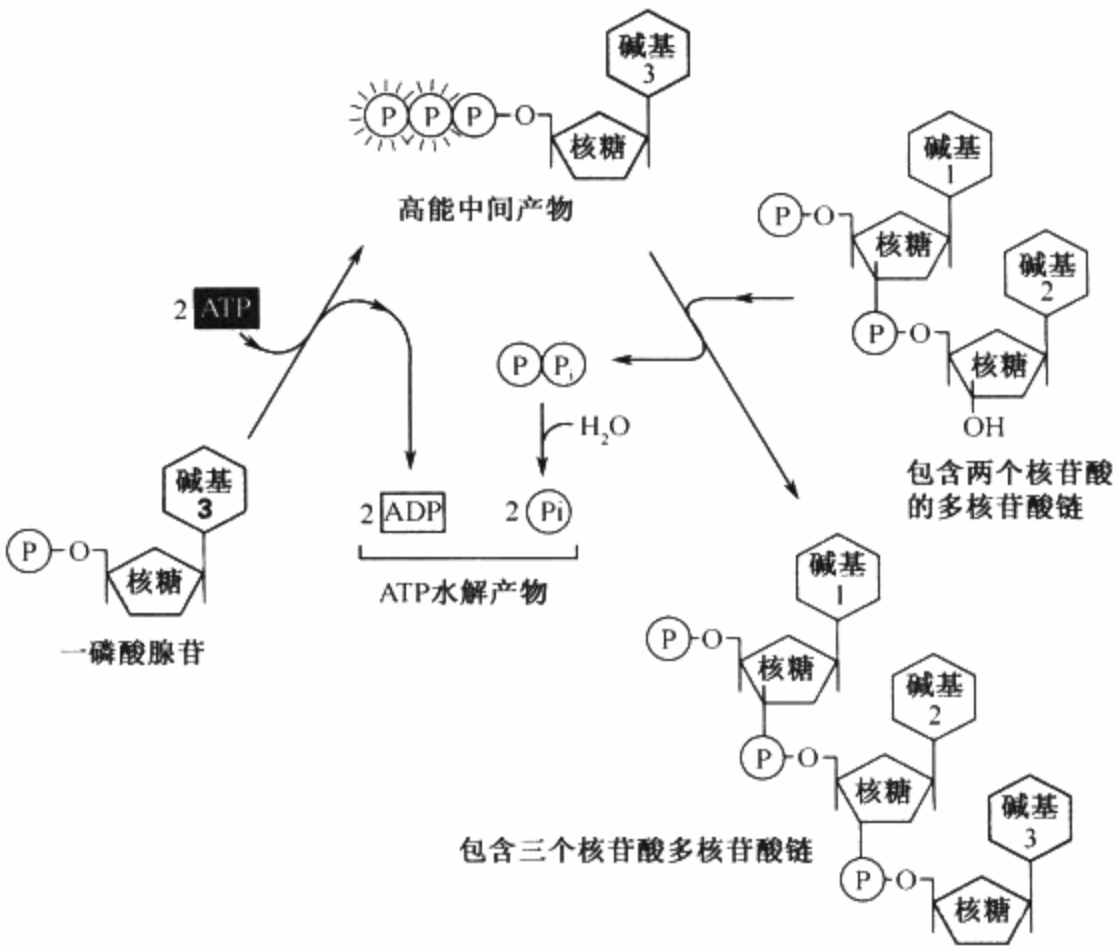


图 2-67　一个多核苷酸，RNA 或 DNA 的合成是一个 ATP 水解推动进行的多步反应过程

在第一步中，核苷单磷酸通过两个 ATP 分子末端磷酸的连续转移反应被活化。形成的高能中间产物——核苷三磷酸，在溶液中游离存在一直到它与生成的 RNA 或 DNA 链的末端进行反应，释放出焦磷酸，后者水解为无机磷酸在能量上极其有利，并且帮助推动整个反应沿核酸合成的方向进行，预知具体细节过程见第 5 章

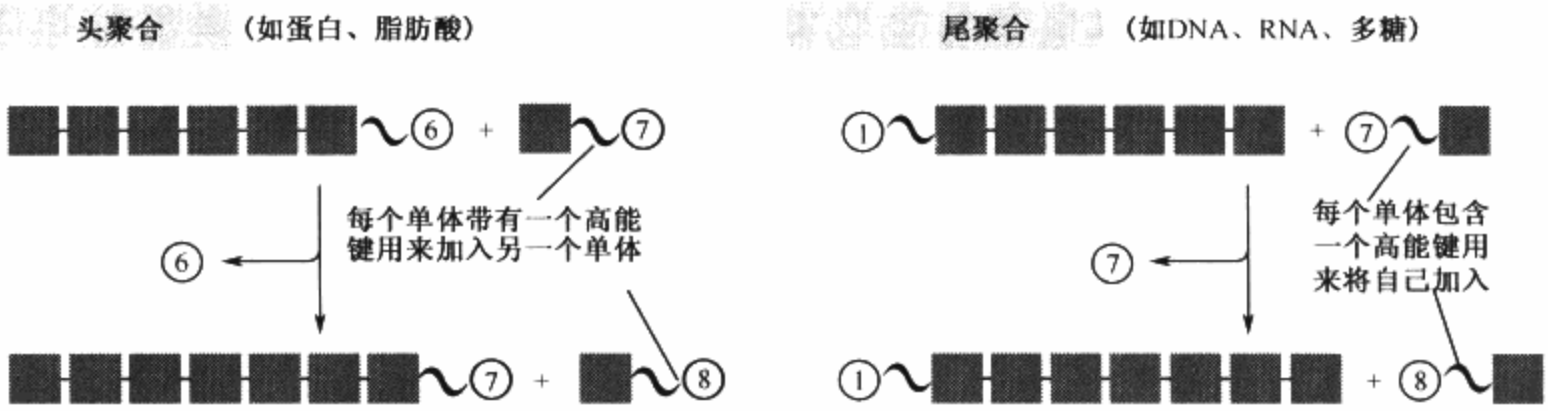


图 2-68　生物聚合反应中活化中间产物的方向性

多聚体的头部生长与对应的尾部生长的比较。如图所示，这两种机制用于合成不同的生物大分子

细胞怎样从食物中获取能量

前面我们刚看到，细胞需要不断提供能量来产生和维持生物有序性以生存下去。这种能量来自于食物分子中的化学键能，因此食物也相当于细胞的燃料一样发挥作用。

糖是特别重要的燃料分子，它们被逐步氧化成 CO_2 和 H_2O（图 2-69）。在这一部分，我们分析糖类物质在分解，或代谢过程中的主要步骤，并揭示它们怎样在动物细胞内产生 ATP、NADH 和其他的活化载体分子。我们主要集中讨论葡萄糖的分解，因为它在大多数动

物的细胞中是主要的产能物质。在植物、真菌和许多细菌中，也存在极其类似的反应途径。其他的分子，如脂肪酸和蛋白质，一旦进入合适的反应途径也能作为能量的来源。

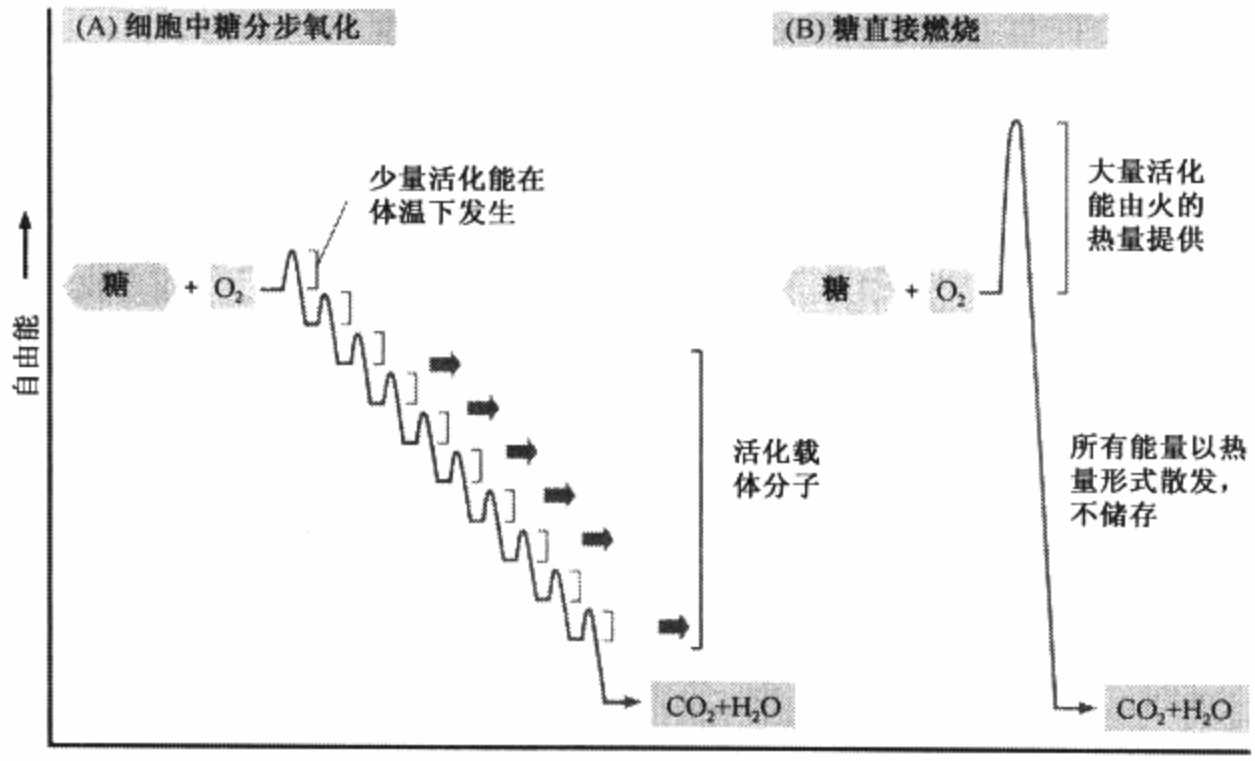


图 2-69　细胞内糖的有控制的逐步氧化过程与普通燃料过程的比较简图

（A）细胞内，酶经过一系列的小步骤催化氧化反应。这个过程中自由能以较方便的一块一块的形式转移到载体分子上，经常是 ATP 和 NADH。在每一步，酶通过降低活化能障碍来控制反应进行，而这个障碍在特定反应发生以前必须逾越。（A）和（B）中释放的总自由能是相同的，但是如果像（B）中那样，糖仅用一步就氧化 CO_2 和 H_2O，那么它就会释放出大大多于可以获取并用于有用目的的能量

食物分子分三个阶段分解产生 ATP

我们所吃的食物中的蛋白质、脂类和多糖必须分解成更小的分子，才能使细胞利用它们，或者作为能量来源或者用作合成其他分子的基本单位。分解过程必须只发生在从外界摄取的食物上，而不能发生在细胞内部的大分子上。因此，食物分子的酶促降解第一阶段是消化，这一过程发生在小肠细胞外或者是细胞内特定的细胞器溶酶体中。（溶酶体外周的膜使得它内部的消化酶与细胞质隔开，如第 13 章所述。）在两种情况下，通过酶的催化作用，食物中巨大的多聚物分子在消化过程中分解成单体分子：蛋白质分解成氨基酸，多糖分解成单糖，脂肪分解成脂肪酸和甘油。消化之后，食物降解生成的有机小分子进入细胞质，在这里开始它们的缓慢的氧化过程。如图 2-70 所示，氧化发生在细胞进一步新陈代谢的两个阶段：第二阶段在细胞质内开始，在主要的能量转换器线粒体内结束；第三阶段则全部局限在线粒体内进行。

在第二阶段，一个称为糖酵解的反应链将每个葡萄糖分子转变成 2 个小分子的丙酮酸。葡萄糖以外的其他单糖，在转变成这条糖分解途径的某个糖中间产物以后，也同样转变成丙酮酸。在形成丙酮酸的过程中，产生了两个活化载体分子——ATP 和 NADH。接着丙酮酸从细胞质进入到线粒体内，在这里每个丙酮酸分子转变成 CO_2 和含有 2 个碳原子的乙酰单位——它们连接到辅酶 A（CoA）上，形成乙酰 CoA，它也是一个载体分子（图 2-62）。来自于脂肪的脂肪酸通过逐步氧化和分解也生成大量的乙酰 CoA，这些脂肪由血液输送，并以脂肪酸的形式进入细胞内，接着进入线粒体产生乙酰 CoA。

在第三阶段，食物分子氧化分解过程完全在线粒体内进行。乙酰 CoA 上的乙酰基团通过

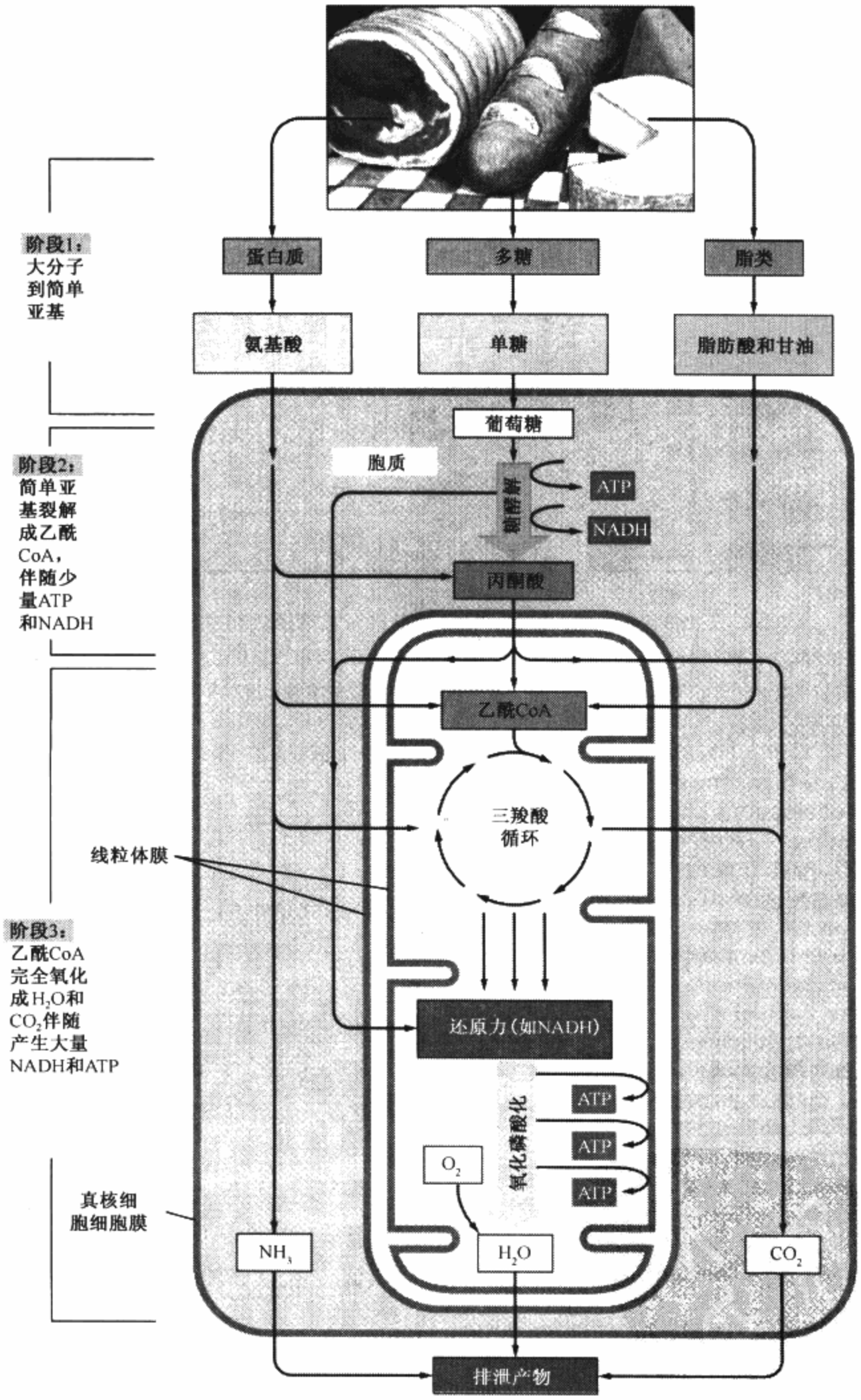


图 2-70　动物细胞内，从食物到无用产物的细胞代谢过程的三个阶段的简图

这一系列的反应产生 ATP，ATP 接着用于细胞内生物合成反应和其他的需能过程。阶段 1 在细胞外进行；阶段 2 主要在细胞质内进行（除了最后一步由丙酮酸转变为乙酰 CoA 上的乙酰基团以外，这一步发生在线粒体体内）；阶段 3 在线粒体内进行

一个高能键与辅酶 A 相连，因而容易转移到其他的分子上。当它转移到含有 4 个碳原子的草酰乙酸上之后，这个乙酰基团进入称为柠檬酸循环的一系列反应过程中。前面我们曾简单提到过，乙酰基团在这些反应中氧化成 CO_2 并生成大量的电子载体 NADH，最后 NADH 携带的高能电子沿着位于线粒体内膜的电子转移链进行传递。在这里，这些电子转移所释放的能量推动生成 ATP 和消耗分子氧（O_2）的过程。正是在最后的这些步骤中，大量氧化释放的能量得到利用并产生在细胞内占大多数的 ATP。

因为线粒体内推动生成 ATP 的能量最终来源于食物分子的氧化分解，所以线粒体中由电子转移推动，使 ADP 磷酸化生成 ATP 的过程也称为氧化磷酸化。第 14 章主要讨论了氧化磷酸化过程中线粒体内膜发生的引人入胜的反应。

通过生成 ATP，糖类和脂肪分解释放的能量以便于在细胞其他地方利用的化学能量块的形式重新分配。一个典型细胞的细胞质内，任何时刻都存在大约 10^9 个 ATP 分子，而且在很多细胞中，所有这些 ATP 每隔 1～2min 就得到更新（也就是消耗竭尽后重新生成并替换）。

总之，葡萄糖和脂肪酸氧化成水和 CO_2 在理论上释放的能量中将近一半可以被获取并用于推动能量上不利的反应 Pi + ADP → ATP。（相反，一个典型的燃料引擎，如汽车引擎，只能将不到 20%的能量转变为有用的功。）剩余的能量被释放，作为热能温暖我们的身体。

糖酵解是生成 ATP 的中心途径

食物分子降解的第二阶段中最重要的过程是葡萄糖在称作**糖酵解**（glycolysis）（“glykos”和“lysis”分别来自希腊语，意思分别为“糖”和“裂解”）的系列反应中分解。糖酵解途径生成 ATP 的过程不需要分子氧（氧气）的参与。它在大多数细胞包括许多厌氧微生物（指那些不需利用氧气生存的微生物）的细胞质中进行。糖酵解途径可能在生命历史的早期，在光合生物释放氧气到空气中以前，就进化产生了。在糖酵解过程中，一个有 6 个碳原子的葡萄糖分子转变成 2 分子的丙酮酸，每个丙酮酸分子含有 3 个碳原子。对于每个葡萄糖分子而言，2 分子的 ATP 水解以提供能量推动初始反应，但是在后面的步骤中，生成了 4 分子的 ATP，因此糖酵解过程结束时，每个葡萄糖分子分解净产生 2 分子的 ATP。

图 2-71 概括了糖酵解途径，专题 2-8 是更详细的反应步骤。糖酵解过程涉及 10 个独立的反应，每个反应产生一个不同的糖中间产物，并且每一步都由不同的酶进行催化。与大多数酶一样，所有这些酶以 ase 结尾，如异构酶（isomerase）和脱氢酶（dehydrogenase），表示它们催化的反应类型。

尽管分子氧不参与糖酵解反应，但是确确实实发生了氧化反应，因为电子被 NAD^+ 从一些来源于葡萄糖分子的碳上移走（产生 NADH）。整个糖酵解过程一步一步地进行，使得氧化产生的能量一部分一部分地释放，这样大多数能量是可以储存到活化载体分子中，而不会使全部的能量以热的形式释放掉（图 2-69）。因此，一些氧化释放的能量推动从 ADP 和 Pi 到 ATP 的直接生成，也有一些能量与高能电子载体 NADH 中的电子一起存在。

糖酵解过程中，每分子葡萄糖分解形成 2 分子 NADH。在需氧生物中（指有氧才能生存的微生物），这些 NADH 分子将它们携带的电子释放到的第 14 章描述的电子传递链中，由 NADH 产生的 NAD^+ 则被糖酵解重新利用（专题 2-8 中第 6 步）。

发酵使得在无氧条件下能够生成 ATP

对于大多数动物和植物细胞而言，糖酵解仅仅是食物分子分解进入第三阶段，也是最后阶段的一个前奏。在这些细胞内，第二阶段最后一步生成的丙酮酸被迅速转移到线粒体内，

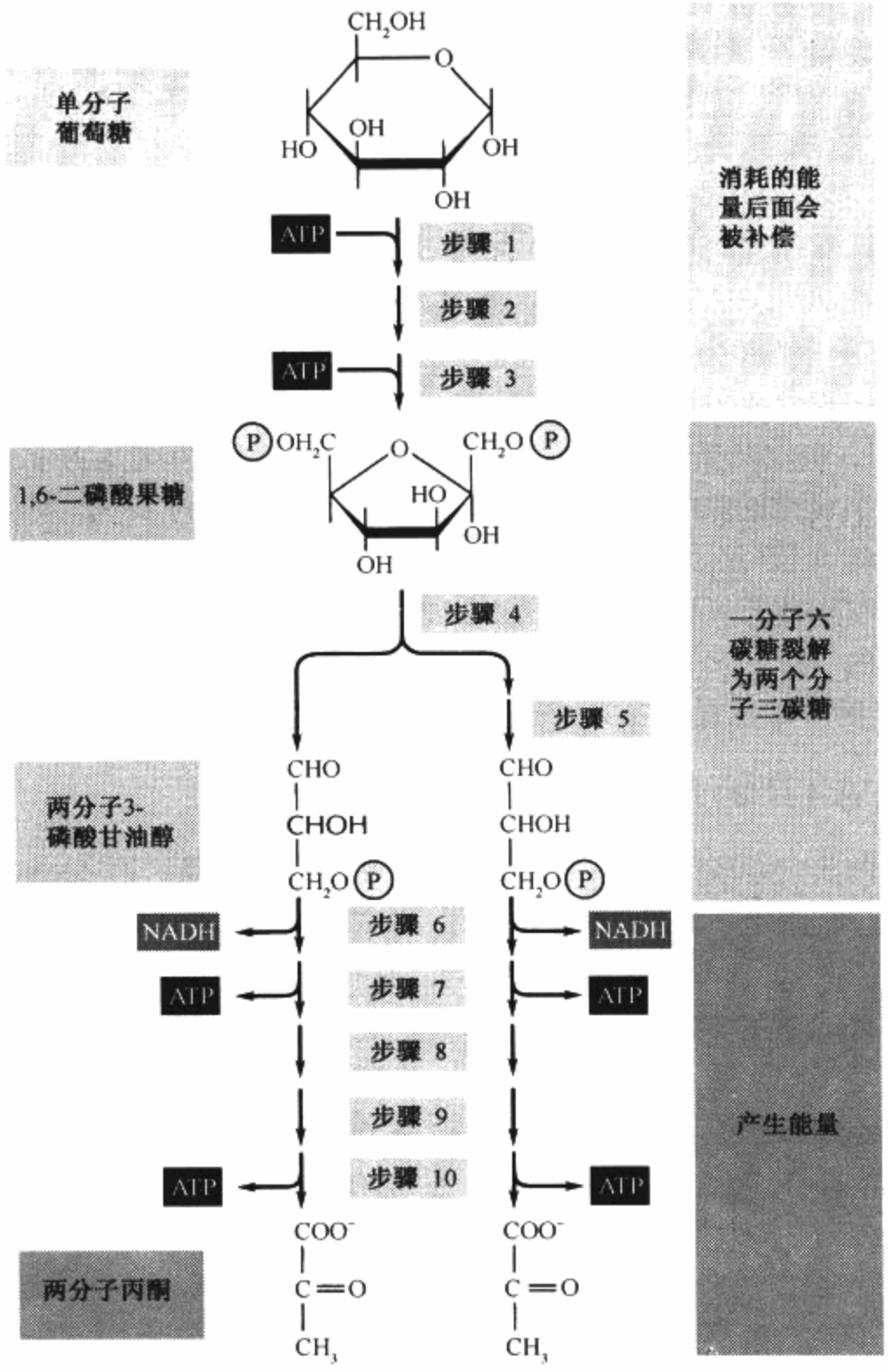


图 2-71　糖酵解途径的概图

图中所示的 10 步反应中每一步都由不同的酶催化，注意第 4 步将 1 个六碳断裂生成 2 个三碳糖，所以这一步后面每一阶段的分子数目增倍。如图所示，从第 6 步开始，糖酵解途径进入产能时期，该时期导致 ATP 和 NADH 分子的净生长（专题 2-8）

在这里丙酮酸转变 CO_2 和乙酰 CoA，而乙酰 CoA 进一步彻底氧化成 CO_2 和 H_2O。

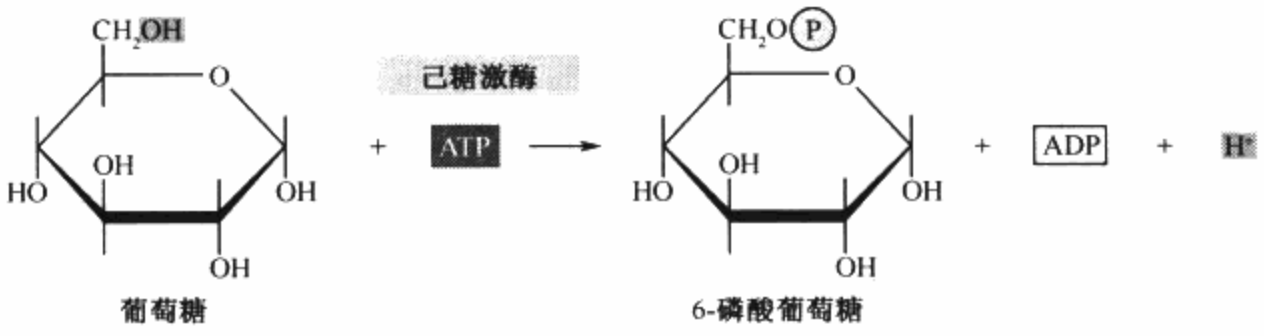
相比之下，对于许多厌氧微生物，它们不利用分子氧，在无氧条件下能够生长和分裂，糖酵解过程是细胞内 ATP 的主要来源。这种情况也适用于某些动物组织，如骨骼肌，它能在分子氧不足的情况下，继续发挥作用。在这些无氧的条件下，丙酮酸和 NADH 电子都存在于细胞质中。丙酮酸转变成分泌到细胞外的产物。例如，在用于酿酒和制作面包的酵母中，转变成乙醇和 CO_2，在肌肉内生成乳酸。在此过程中，NADH 释放它的电子又转变成 NAD^+。这种 NAD^+ 再生的过程对于维持糖酵解途径正常运行是必需的（图 2-72）。

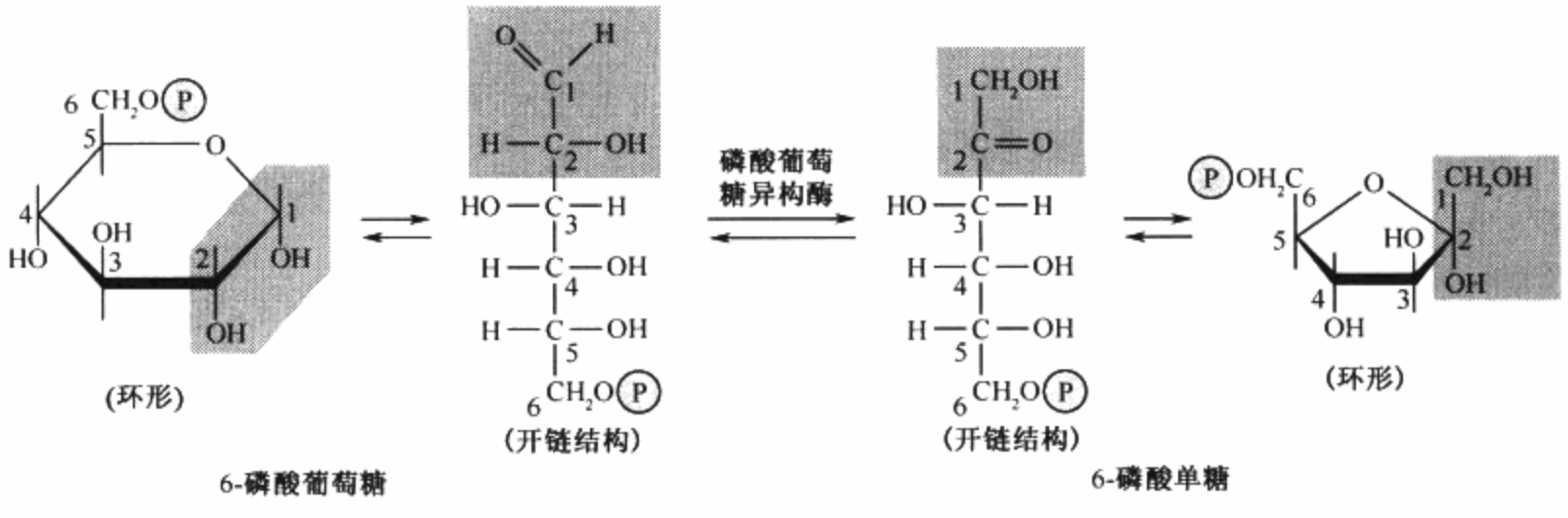
这种无氧条件下产生能量的途径称为**发酵**（fermentation）。早期有关在酵母中进行的，且在商业上很重要的发酵反应的研究给了生物化学家灵感。19 世纪的研究工作导致研究者在

专题 2-8　糖酵解途径中 10 个步骤的详细内容

每一步反应中发生变化的分子用蓝色阴影表示，而黄色方框中是催化反应进行的酶。

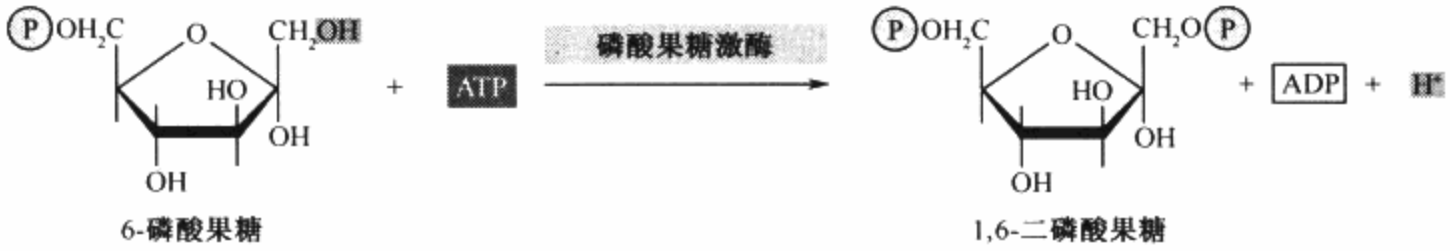
第 1 步：葡萄糖被 ATP 磷酸化，生成糖磷酸。磷酸所带的负电荷阻止糖磷酸通过细胞质膜，将葡萄糖限制在细胞内。



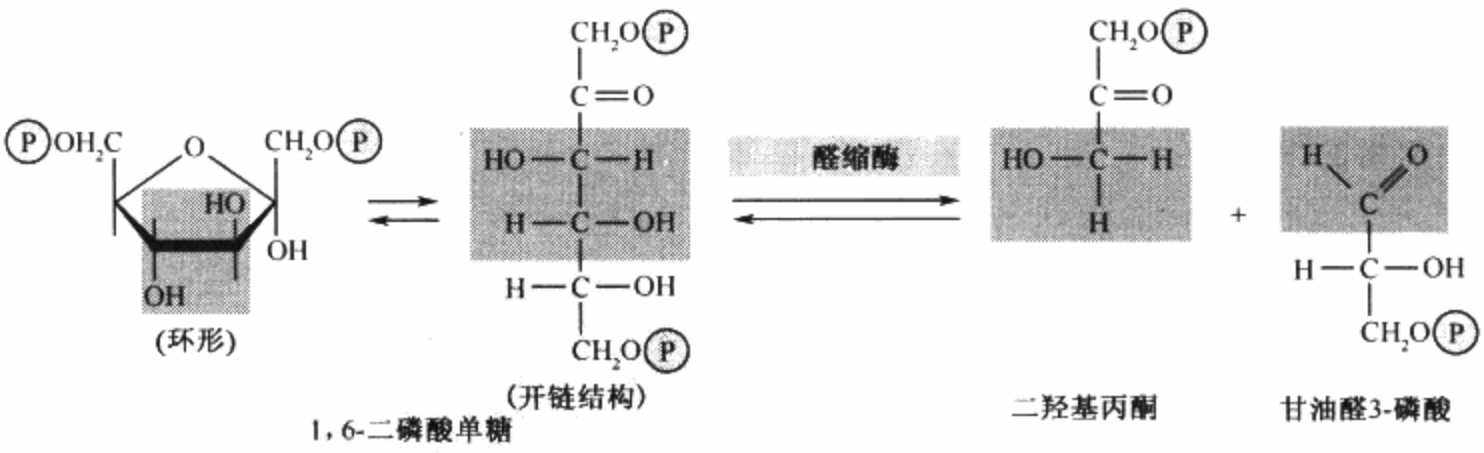


第 2 步：化学结构的一种极易可逆的重排将羰基氧从 C1 转移到 C2（异构化），从而由醛糖生成酮糖（图 2-4）。

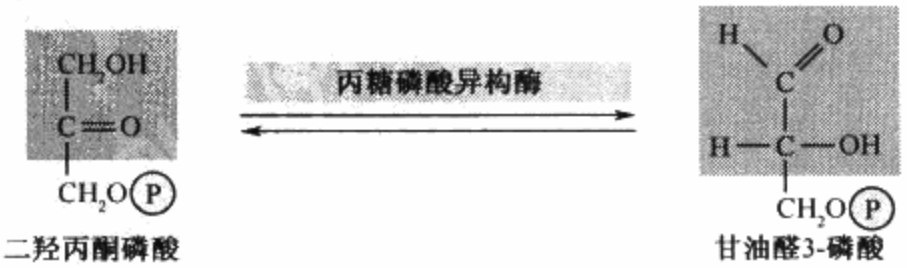
第 3 步：C1 位上新出现的羟基被 ATP 磷酸化，从而为生成 2 个三碳糖磷酸作好准备。通过对磷酸果糖激酶的调节，糖类物质进入糖酵解途径受到控制。

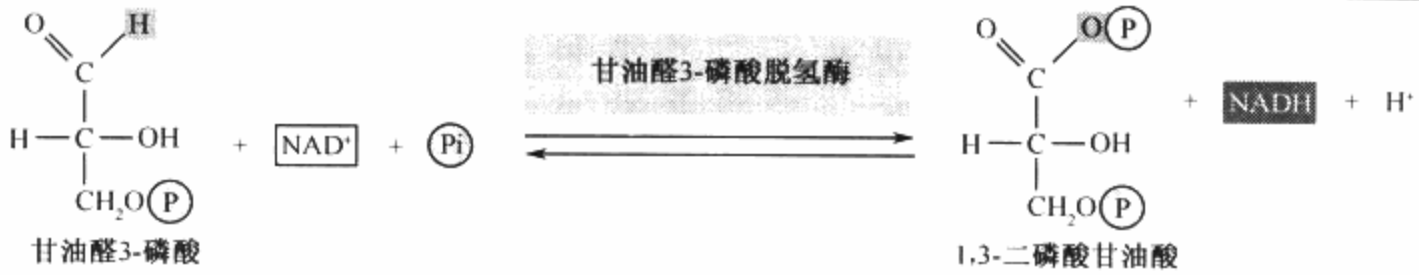


第 4 步：六碳糖断裂生成 2 个三碳分子，只有甘油醛 3-磷酸能立即沿着糖酵解途径继续反应。

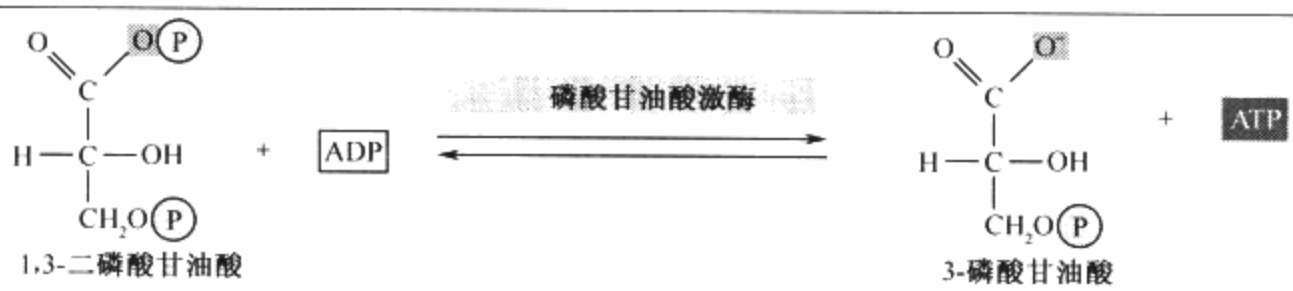


第 5 步：第 4 步中的另外一个产物——二羟丙酮磷酸，异构化生成甘油醛 3-磷酸。



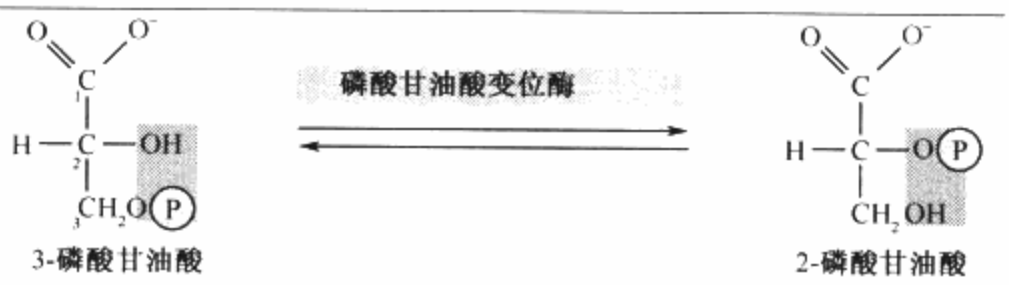


第 6 步：2 分子的甘油醛 3-磷酸被氧化，由于形成 NADH 和一个新的高能酐磷酸连接，糖酵解途径进入产能阶段。

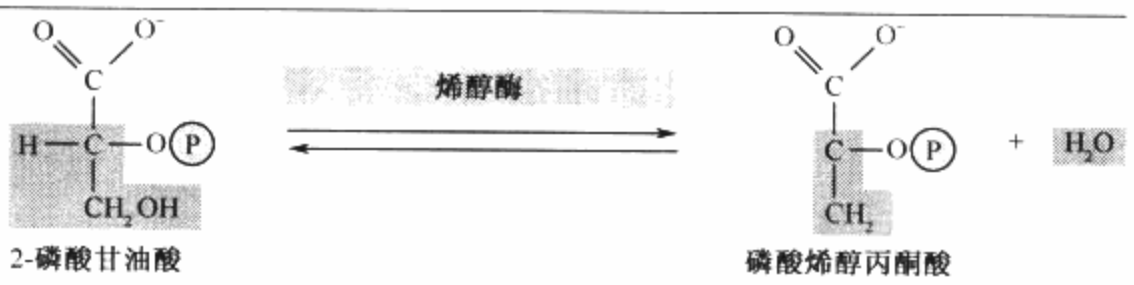


第 7 步：第 6 步生成的高能磷酸键转移到 ADP 上生成 ATP。

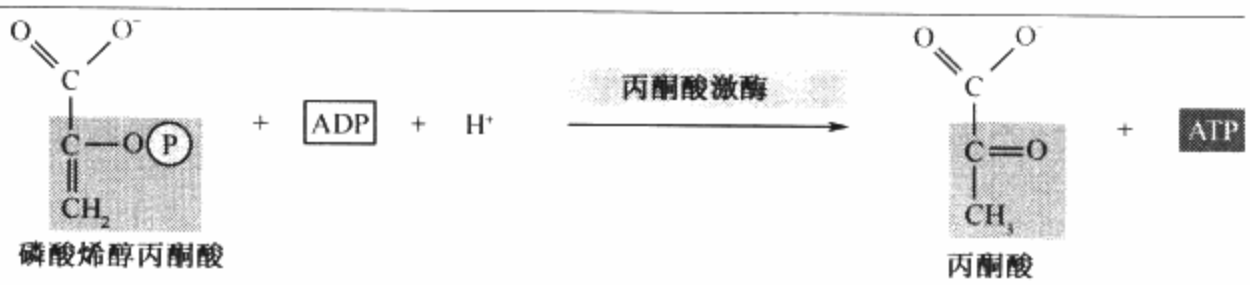
第 8 步：3-磷酸甘油酸中剩余的一个磷酯连接具有相对较低的水解自由能，从 C3 转移到 C2，生成 2-磷酸甘油酸。



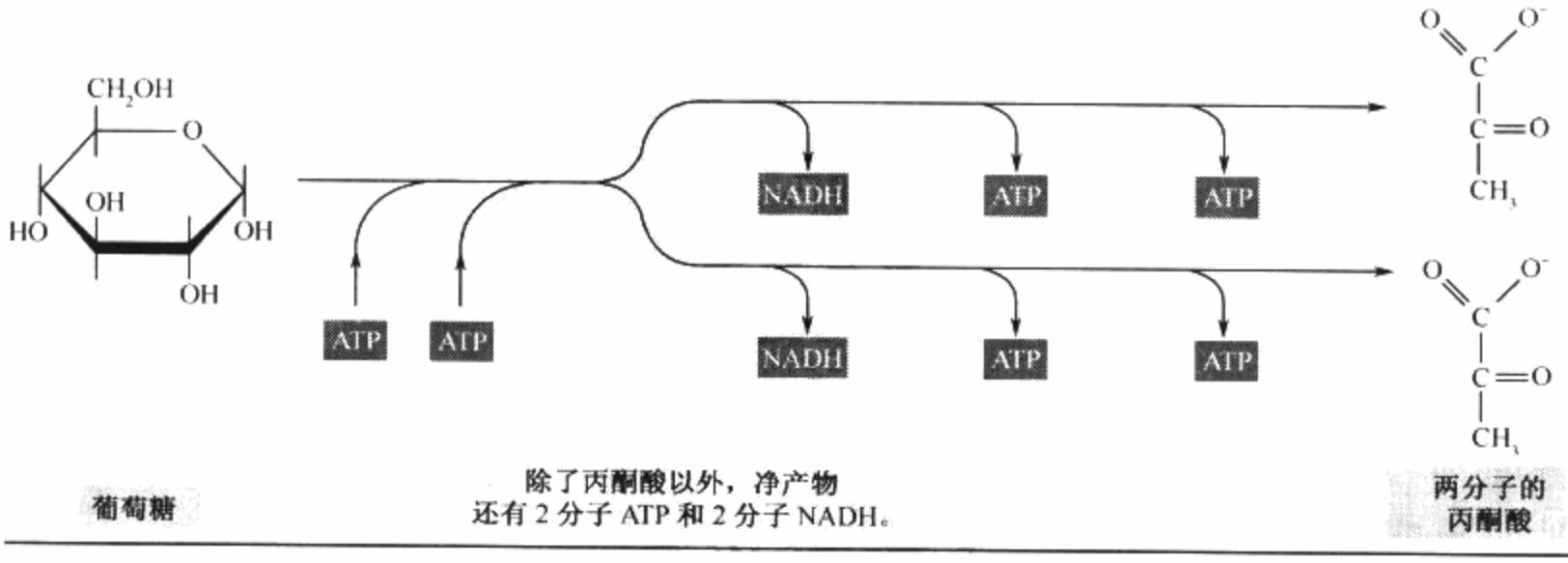
第 9 步：从 2-磷酸甘油酸移走一分子水生成一个高能烯醇磷酸连接。



第 10 步：第 9 步生成的高能磷酸基团转移到 ADP 生成 ATP，完成糖酵解过程。



糖酵解的净结果



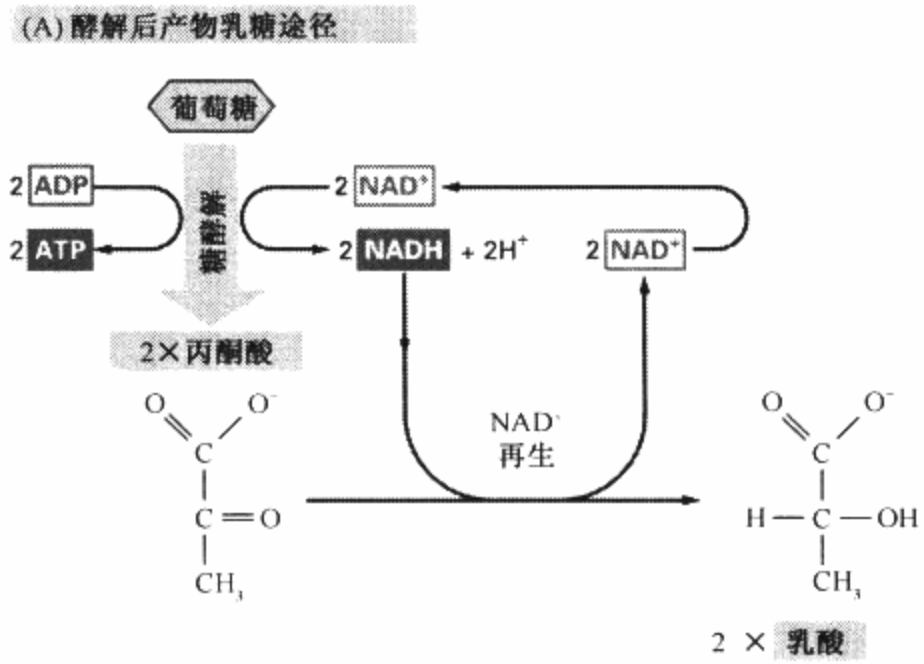


图 2-72　丙酮酸无氧分解的两条途径

（A）当氧气供应量不足时，例如，在剧烈收缩的肌肉中，糖酵解产生的丙酮酸转变成乳酸，如图所示。这个反应再生出糖酵解途径第 6 步所消耗的 NAD^+，但是整个途径产生的能量比完全氧化产生的能量少得多。（B）在一些无氧条件下也能生长的有机体。例如，在酵母中，丙酮酸经由乙醛转变成二氧化碳和乙醇。同样，该途径从 NADH 再生出糖酵解途径正常运行所需的 NAD^+。（A）和（B）都是发酵的例子

1896 年令人惊奇地认识到这些过程可以在活生物体外进行研究，也就是在细胞抽提物内进行研究。这个革命性的发现最终使研究者能够分解并研究发酵过程中每个单独的反应。1930 年拼装完成完整的糖酵解途径，这是生物化学的一大成就，接着很快研究者就认识到细胞反应过程中 ATP 的中心作用。因此，本章讨论的大多数基本概念已经了解了 50 多年。

糖酵解过程证实了酶是如何将氧化放能与能量储存偶联起来的

前面我们已使用一个“叶轮”的类比来解释细胞如何利用酶将一个能量上不利的反应与另一个能量上有利的反应偶联起来，从而获取有机分子氧化产生的有用能量（图 2-56）。在我

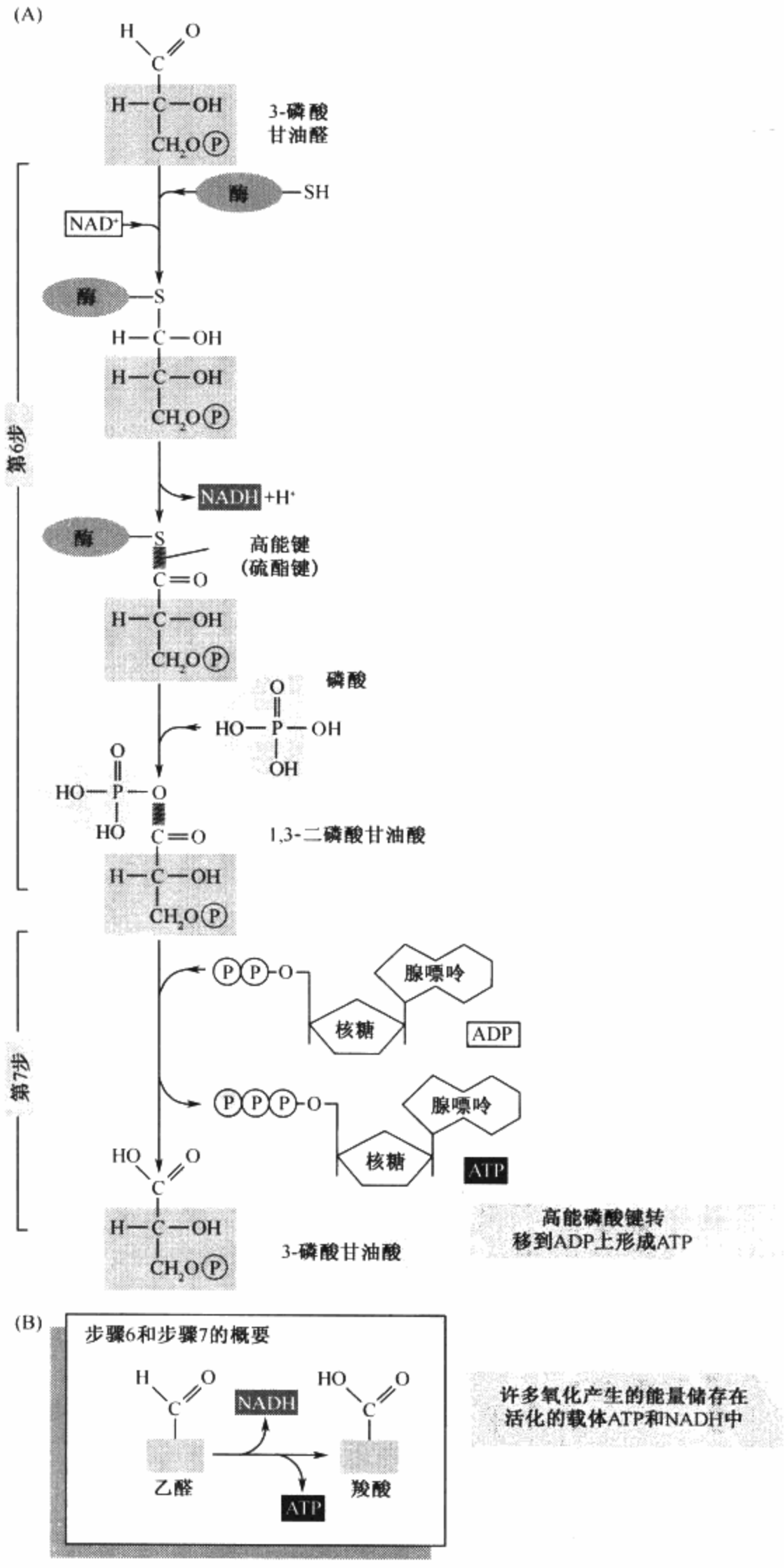


图 2-73　糖酵解途径的第 6、7 两步中的能量储存

在这些反应步骤中，醛氧化为羧酸的过程与生成 ATP 和 NADH 的过程相偶联。（A）第 6 步开始于底物（甘油醛 3-磷酸）和暴露在酶表面（甘油醛 3-磷酸脱氧酶）的一个—SH 基团之间形成共价键。然后这个酶催化将氢（是一个氢负离子，即一个质子加 2 个电子）从大量存在的甘油醛 3-磷酸转移到一个 NAD^+ 分子上。在这个氧化反应过程中释放的能量部分用于生成一分子 NADH，部分用于将起始时酶与底物之间形成的共价键转变成一个高能硫酯键（用红色表示），接着，一分子无机磷酸取代酶上面的这个高能键，从而生成一个高能糖磷酸键（红色表示）。到此，酶不仅将能量储存于 NADH，而且将能量上有利的使醛氧化为羧酸的反应偶联到生成高能磷酸键的能量上不利的反应上。第 2 个反应由第 1 个反应推动，因而就像图 2-56 中的“叶轮”偶联机制一样发挥作用。在第 7 步反应中，高能糖-磷酸中间产物刚刚形成，1,3-二磷酸甘油酸就结合到另一个酶磷酸甘油激酶上，活性磷酸转移到 ADP 上，形成一分子 ATP，并将一个自由的羧酸基团保留在被氧化的糖上面。（B）对第 6、7 两步反应中发生的整个化学变化的概括

们的类比中，酶相当于叶轮，为了确切地说明偶联反应是如何发生的，现在我们回到前面讲的糖酵解途径的一个步骤中去。

糖酵解过程中的两个中间反应（第 6 步和第 7 步），将含有 3 个碳原子的糖中间产物甘油醛 3-磷酸（一个醛）转变成了一磷酸甘油酸（一个羧酸）。这个过程使得一个醛基在两步进行的反应中氧化成羧基。整个过程释放了足够的自由能将 ADP 分子转变成 ATP，也将两个电子从醛基上转移到 NAD^+ 形成 NADH，而且释放了足够的热能到环境中，使得整个反应在能量上有利（整个反应的 ΔG 是 −3.0kcal/mol）。

图 2-73 列举了实现这种显著转变的反应途径。这些化学反应由两种酶催化控制，而两种糖中间物结合在这些酶上。第一个酶（甘油醛-3-磷酸脱氢酶）通过酶分子上具有活性的—SH基团，形成一个短暂的共价键，结合到醛基上。尽管处于结合状态，它也能催化这个醛基的氧化。氧化产生的高能酶-底物共价键，接着被一个无机磷酸离子取代，生成一个高能糖-磷酸中间产物，因而这个中间产物可以从酶分子上释放出来。接着结合到第二个酶（磷酸甘油激酶）上。这个酶催化刚生成的高能产物磷酸转移到 ADP，生成 ATP，并完成将醛基氧化为羧酸的整个过程（图 2-73）。

我们较详细地介绍了整个特别的氧化过程，因为它提供一个明显的由酶介导的能量储存的实例（图 2-74）。这些反应（第 6 步和第 7 步），是糖酵解途径中仅有的直接从无机磷生成高能磷酸键的反应。因此，它们导致每个葡萄糖分子净生成 2 个 ATP 分子和 2 个 NADH 分子（专题 2-8）。

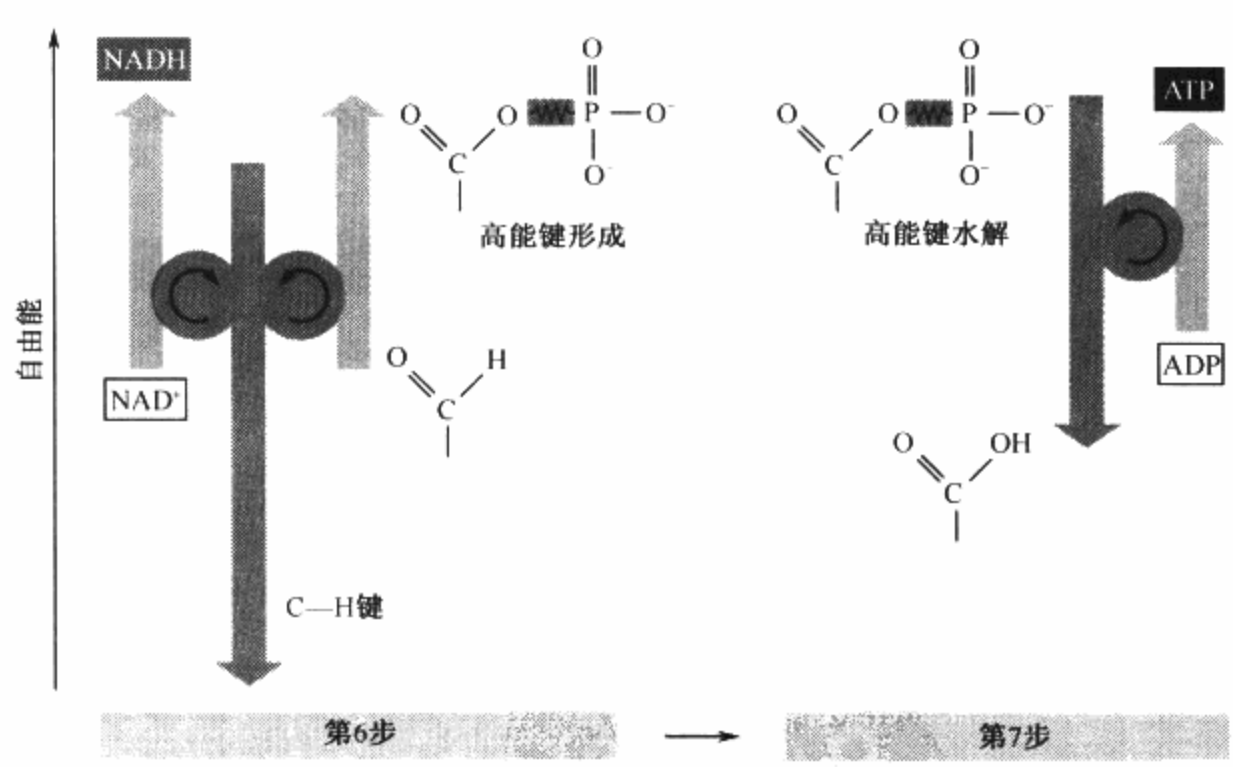


图 2-74　糖酵解途径中第 6 步和第 7 步反应中形成 NADH 和 ATP 的偶联反应概览

C—H 键氧化的能量推动生成 NADH 和一个高能磷酸键，这个高能磷酸键断裂推动生成 ATP

正如我们刚看到的，当生成的反应中间产物具有比 ATP 更高的高能磷酸键时，ATP 很容易由 ADP 生成。磷酸键可以通过比较每个键水解断裂时标准自由能变化（ΔG^0），按能量进行排序。图 2-75 比较了 ATP 中的高能磷酸键和其他磷酸键，其中有一些在糖酵解途径中产生。

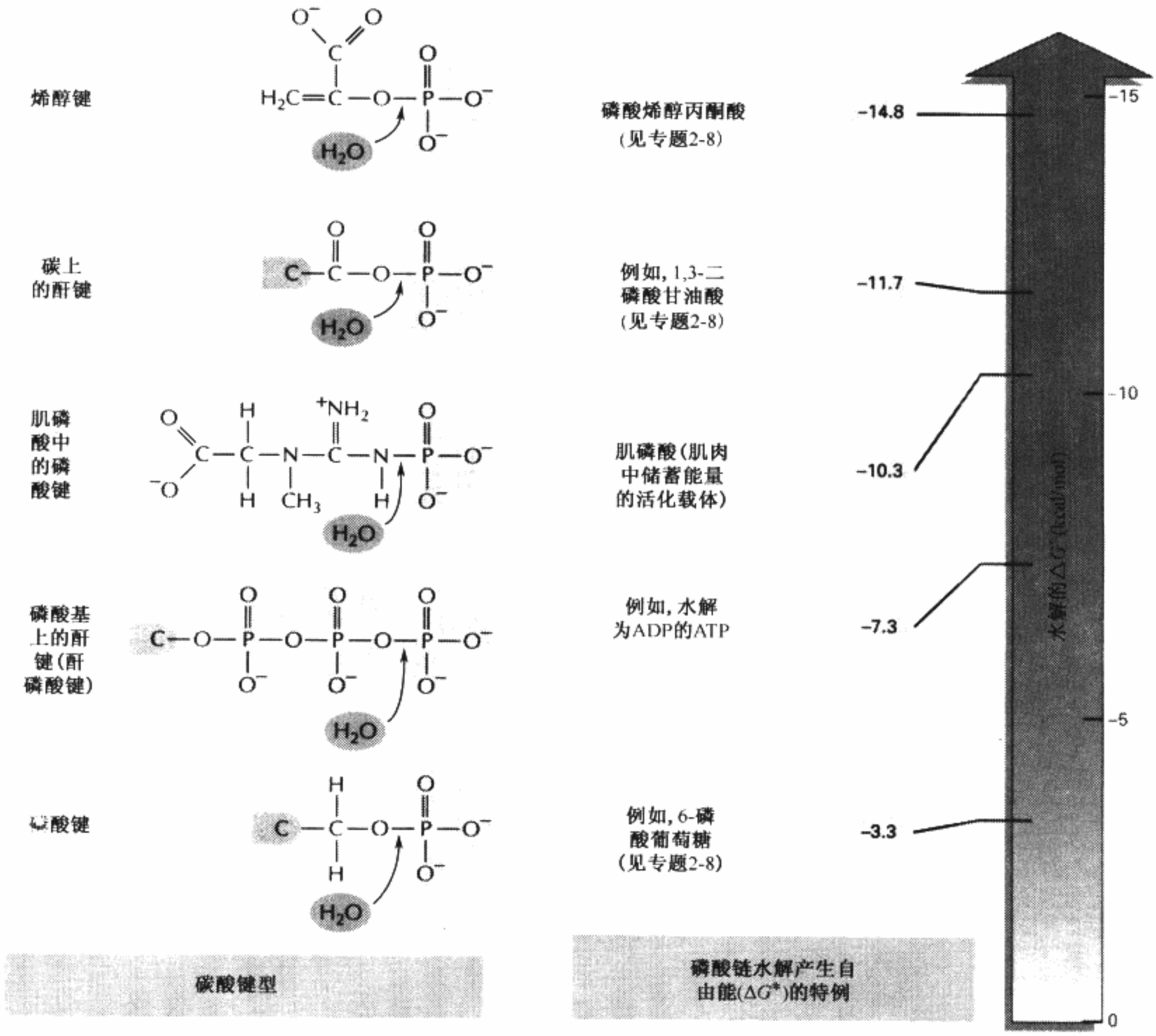


图 2-75　一些磷酸键的能量

如果水解分子 1 中的磷酸键的标准自由能变化比水解分子 2 的更负，那么将此磷酸基团从分子 1 转移到分子 2 的反应在能量上就有利。因此，例如，一个磷酸基团很容易从 1,3-二磷酸甘油酸转移到 ADP，生成 ATP。注意水解反应可看到将磷酸基团转移到水分子

糖和脂肪都在线粒体分解为乙酰 CoA

现在我们来看代谢的第 3 个阶段，也就是需要大量分子氧（氧气）的阶段。由于研究者们认为地球在 10 亿～20 亿年以前形成了含有氧气的大气层，然而我们知道大量的生命形态已在地球上存在了 35 亿年，所以我们接着讨论的在反应中利用 O_2 被认为是一个相对较晚的事件。相比之下，图 2-73 中产生 ATP 的机制不需要氧气，并且与这对偶联反应相近的机制可能在地球生命历史早期就出现了。

在需氧代谢中，糖酵解途径产生的丙酮酸被一个由三种酶组成的巨大复合体——丙酮酸脱氢酶复合体，迅速脱氢。丙酮酸脱氢产物包括 1 分子 CO_2（无用产物）、1 分子 NADH 和乙酰 CoA。该三酶复合体位于真核细胞线粒体内；它的结构和作用模式如图 2-76 所示。

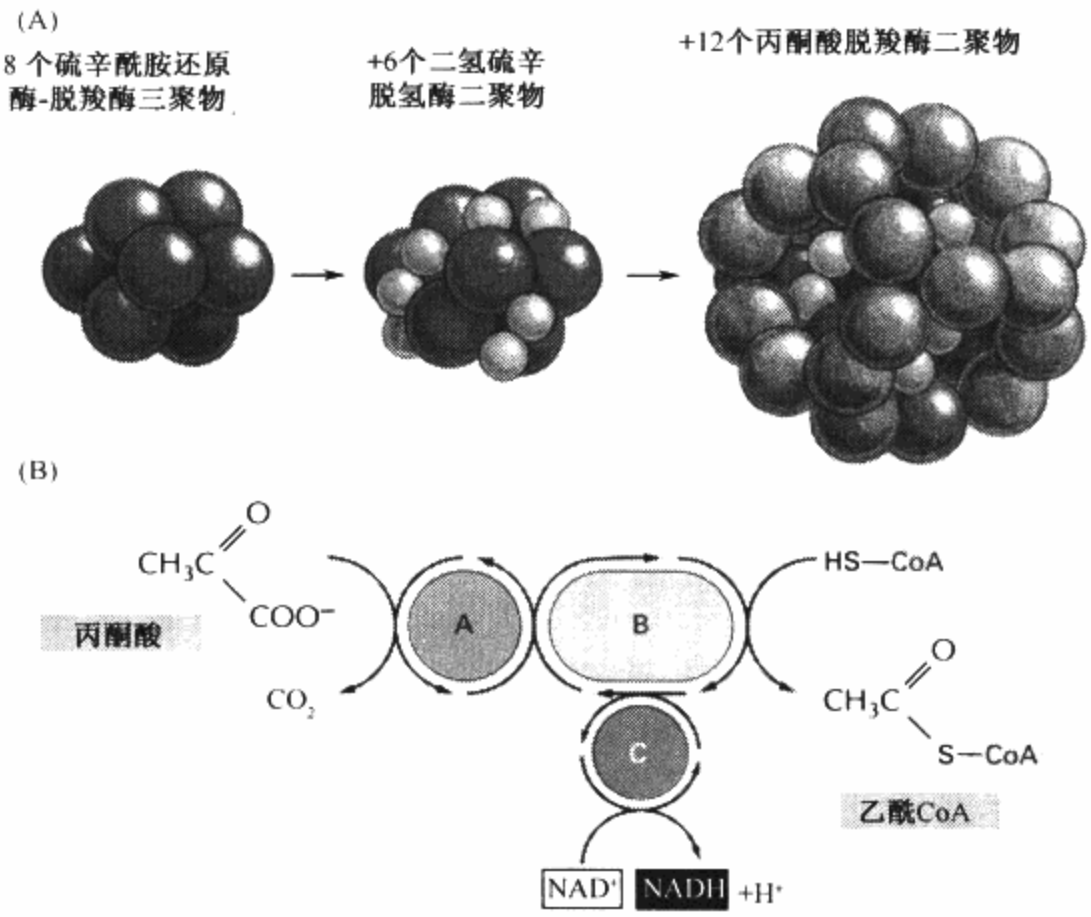


图 2-76　丙酮酸氧化成乙酰 CoA 和 CO_2

（A）丙酮酸脱氢酶复合体的结构，它含有 60 条多肽链。这是一个大的多酶复合体的例子，在此复合体内反应中间产物直接从一个酶传递到另一个酶。在真核细胞中它位于线粒体。（B）丙酮酸脱氢酶复合体发生的反应。该复合体在线粒体基质中将丙酮酸转变成乙酰 CoA，这个反应过程也生成 NADH。A、B、C 分别指三种酶：丙酮酸脱羧酶、硫辛酰胺还原酶脱羧酶和二氢硫辛脱氢酶。（A）中用图表示这些酶，它们的活性联系也如图所示

降解来源于脂肪的脂肪酸的酶同样在线粒体内生成乙酰 CoA。每个脂肪酸分子（如活化的脂酰 CoA 分子）经过一个循环反应完全分解，整个循环反应每次从羧基末端去掉 2 个碳并且每轮循环生成 1 分子乙酰 CoA。这个过程中还生成 1 分子 NADH 和 1 分子 $FADH_2$（图 2-77）。

糖和脂肪为大多数非光合生物，包括人类，提供了主要的能量来源。然而，从这两种类型食物的氧化中，可获得的有用能量储存在乙酰 CoA 分子中，它们经由上面所述的两种反应产生。由于乙酰 CoA 中的乙酰基团在柠檬酸循环中被氧化成 CO_2 和 H_2O，因此在需氧生物的能量代谢中，柠檬酸循环发挥着中心作用。在真核生物中，这些反应都发生在线粒体内，丙酮酸和脂肪酸都进入这个细胞器生成乙酰 CoA（图 2-78）。因此，应当毫不奇怪动物细胞大部分 ATP 的产生都发生在线粒体内。相反，厌氧细菌在唯一的区室细胞质内进行所有的反应，而且在这些细胞内柠檬酸循环正是在细胞质中进行的。

柠檬酸循环使乙酰 CoA 氧化成 CO_2，生成 NADH

19 世纪生物学家注意到没有空气时（厌氧条件），细胞产生乳酸（如在肌肉内）或者乙醇（如在酵母内），而存在空气时（需氧条件），细胞消耗 O_2 并生成 CO_2 和 H_2O。为了阐明需氧代谢途径，所做的大量工作最后集中在丙酮酸氧化上面，从而使得 1937 年发现了**柠檬酸循环**（citric acid cycle），也称为三羧酸循环或者 Kreb's 循环。柠檬酸循环氧化的碳化合物在大多数细胞内约占 2/3，它的主要终产物是 CO_2 和以 NADH 形式存在的高能电子。CO_2 作为无用产物释放，而 NADH 中的高能电子转移到结合在膜上的电子传递链，最终结合氧气生成 H_2O。尽管柠檬酸循环本身并不利用 O_2，但它需要 O_2 以不断向前进行，因为没有其他有效

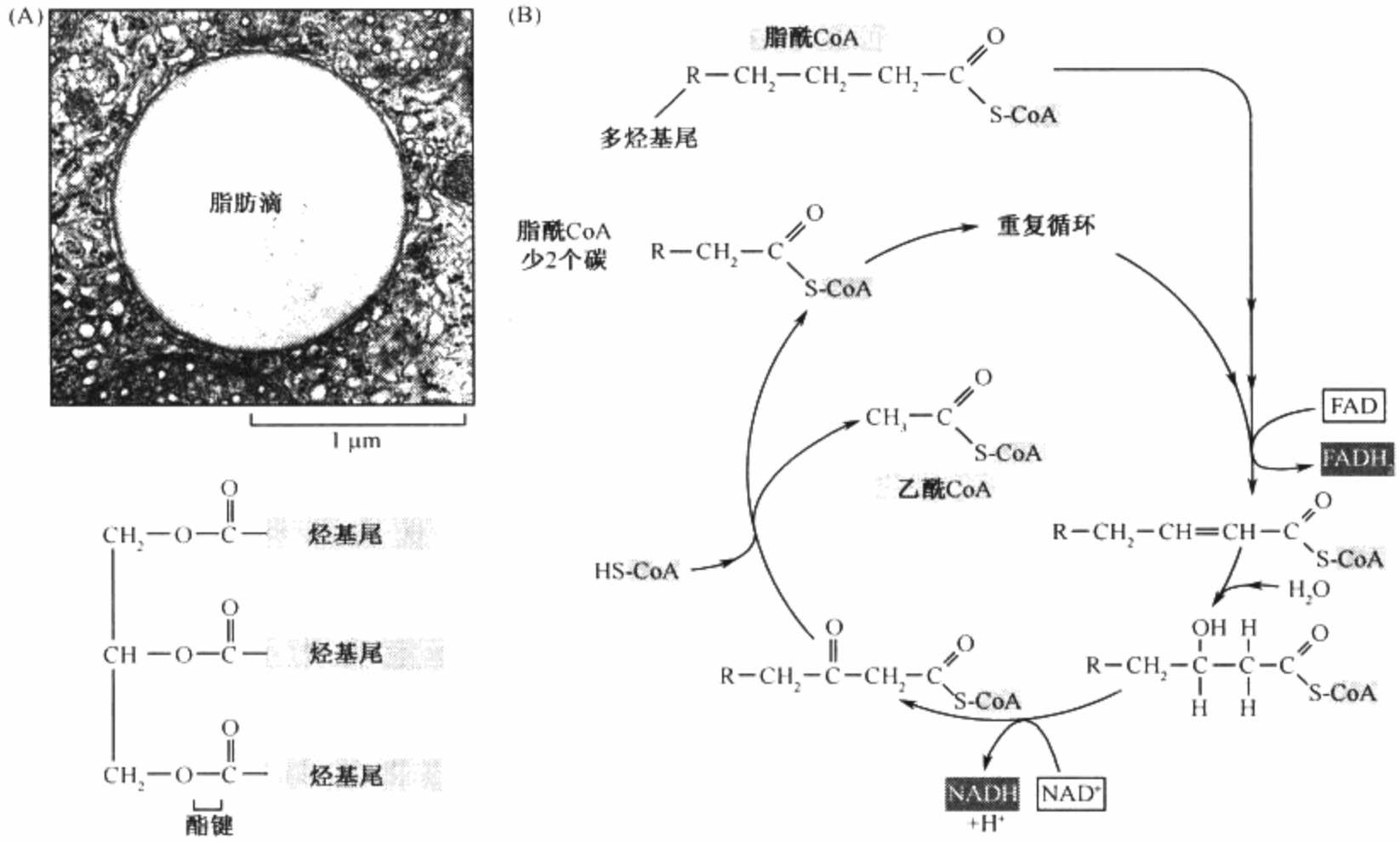


图 2-77　脂肪酸氧化成乙酰 CoA

（A）细胞质（上面）内脂肪滴的电镜图以及脂肪的结构（下面）。脂肪是三酰甘油。这里用绿色表示与三个脂肪酸用酯键相连的甘油。脂肪不溶于水，储存在特化的脂肪细胞（称为 dipocyte）内，形成巨大的脂肪滴。（B）脂肪酸氧化循环。这个循环由线粒体内的 4 种酶催化。每轮循环使脂肪酸链减少 2 个碳原子（用红色表示），并生成 1 分子乙酰 CoA. 1 分子 NADH 和 1 分子 $FADH_2$。$FADH_2$的结构如图 2-80（B）所示［（A）经 Daniel S. Frednd 许可］

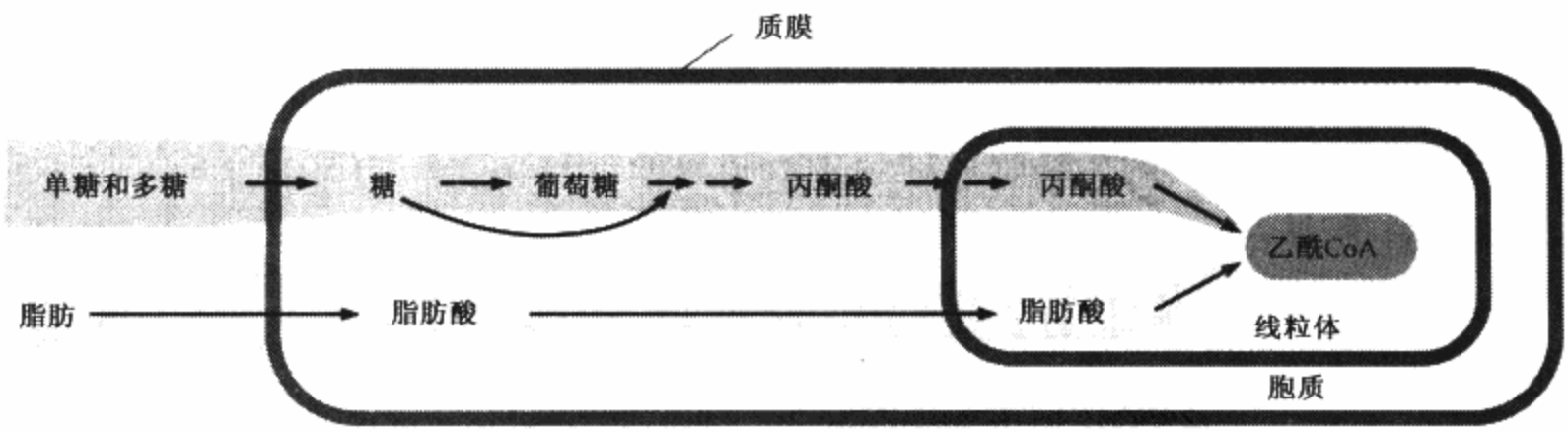


图 2-78　从糖和脂肪生成乙酰 CoA 的途径

真核细胞内的线粒体是两种主要的食物分子产生乙酰 CoA 的地方。因此，正是在线粒体发生大多数的细胞氧化反应，并且生成大多数的 ATP

途径可以使 NADH 释放它的电子，因此 NAD^+ 的再生是循环持续进行所必需的。

柠檬酸循环在真核生物内主要在线粒体内进行，它使乙酰 CoA 中乙酰基团的碳原子完全氧化，转变为 CO_2。但是，乙酰基团并不是直接氧化的。相反，这个基团从乙酰 CoA 转移到一个更大的四碳分子——草酰乙酸，并形成一个六碳的三羧酸——柠檬酸，后面的反应循环就是根据它命名的。柠檬酸分子接着被缓慢氧化，使得氧化释放的能量可以被利用，并产生富含能量的活化载体分子。串联的 8 个反应组成一个循环，因为最后草酰乙酸再生并进入新

一轮循环当中，如图 2-79 所示。

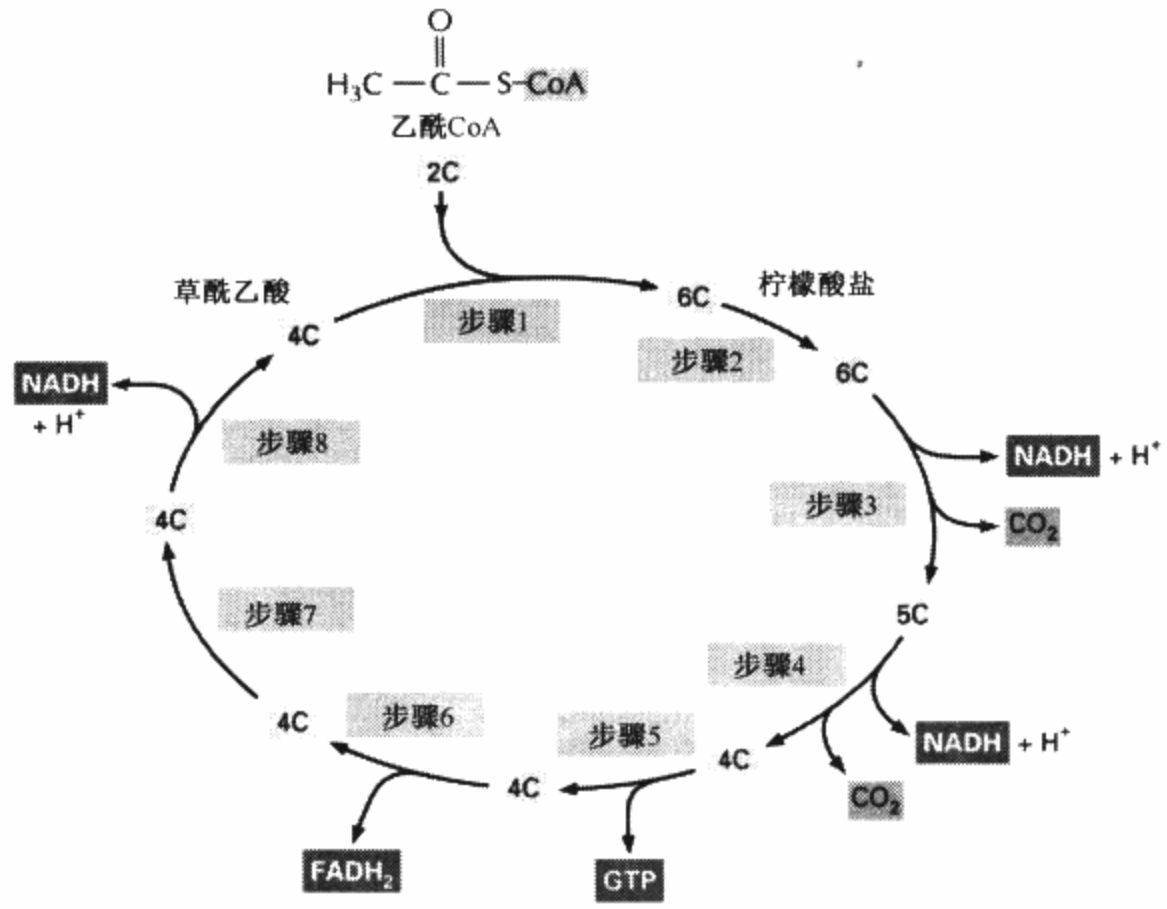


图 2-79　柠檬酸循环的简图

乙酰 CoA 和草酰乙酸反应生成柠檬酸（citric acid），从而启动柠檬酸循环。每一轮循环里，生成 2 分子 CO_2 作为无用产物，再产生 3 分子 NADH、1 分子 GTP 和 1 分子 $FADH_2$。每个中间产物的碳原子数用黄色方格表示，欲知细节见专题 2-9

到目前为止，我们仅仅讨论了三种活化载体分子中的一种——NAD^+-NADH 对（图 2-60）。除了分子 NADH 以外，每轮循环还从 FAD 产生 1 分子 **$FADH_2$**（还原型黄素核苷二磷酸）以及从 GDP 产生 1 分子 **GTP**（鸟嘌呤核苷三磷酸）。这两种活化载体分子的结构见图 2-80。GTP 与 ATP 类似，每轮循环中，GTP 末端磷酸基团转移到 ADP 生成 1 分子 ATP。与 NADH 一样，$FADH_2$ 是高能电子和氢的载体。正如我们刚提到的，储存在 NADH 和 $FADH_2$ 中的易转移的高能电子的能量，通过氧化磷酸化过程，将接着用于 ATP 合成，它是食物氧化代谢过程中唯一一步直接需要空气中的氧气（O_2）进行的反应。

在专题 2-9 中，我们可以看到完整的柠檬酸循环途径。为了使进入柠檬酸循环的乙酰基团生成 CO_2 所必需的额外氧原子，不是由分子氧而是由 H_2O 提供。正如专题 2-9 中所示，每轮循环中 3 分子水裂解，其中一些氧原子用于合成 CO_2。

除丙酮酸和脂肪酸以外，一些氨基酸通过细胞质进入线粒体，在这里转变成乙酰 CoA 或者柠檬酸循环的一个中间产物。因此，在真核细胞内线粒体居中心地位，所有产能过程，无论它们从糖、脂肪还是蛋白质开始，都导向这个中心。

柠檬酸循环也可以作为许多重要生物合成反应的起点，为它们提供至关重要的含碳中间产物，如草酰乙酸和 α-酮戊二酸。代谢产生的这些物质中有些从线粒体转运回细胞质，在这里它们作为许多重要的分子，如氨基酸合成的前体发挥作用。

电子转移推动细胞内大多数 ATP 的合成

正是在食物分子降解的最后一步，它所含化学能的大部分被释放。在这个最后的反应过

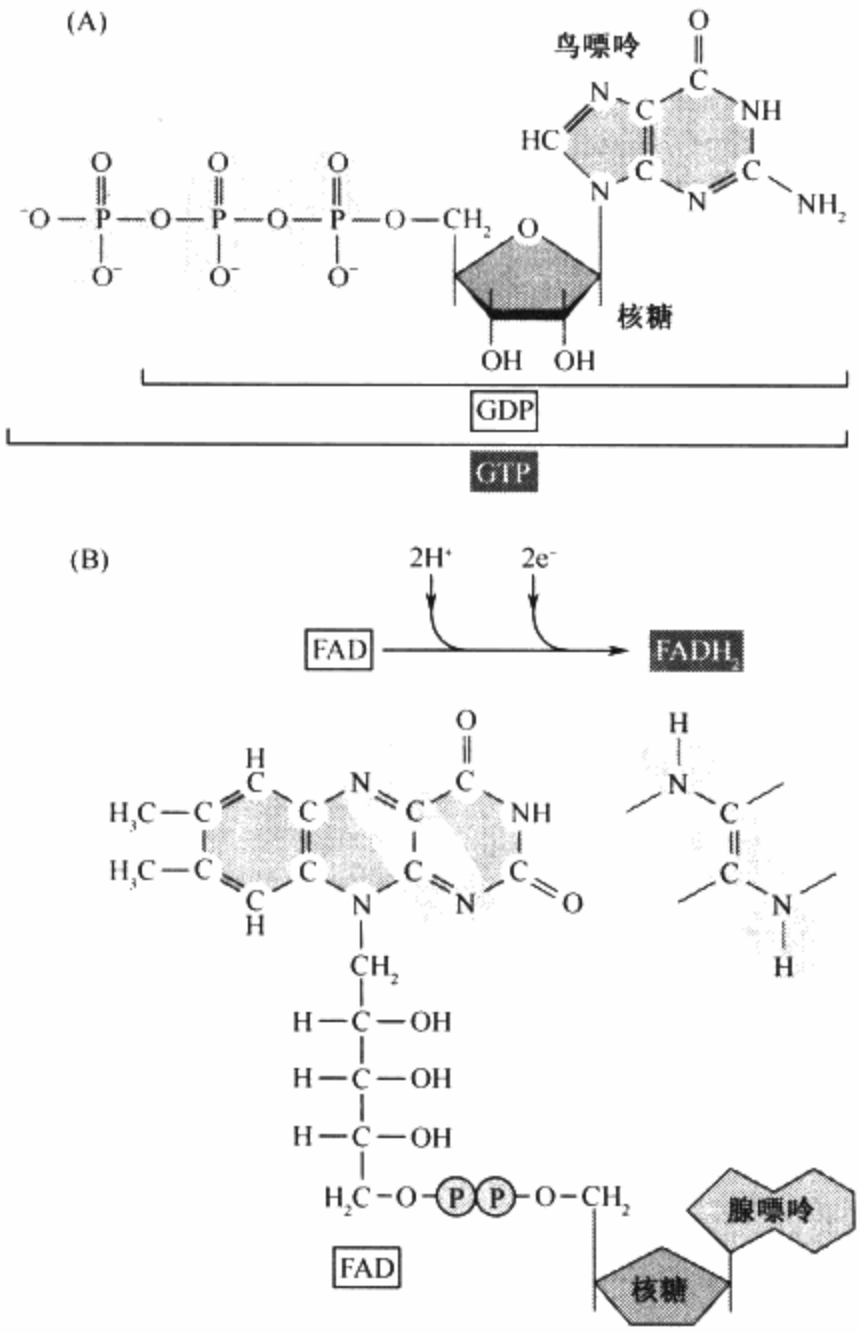


图 2-80　GTP 和 FADH 的结构图

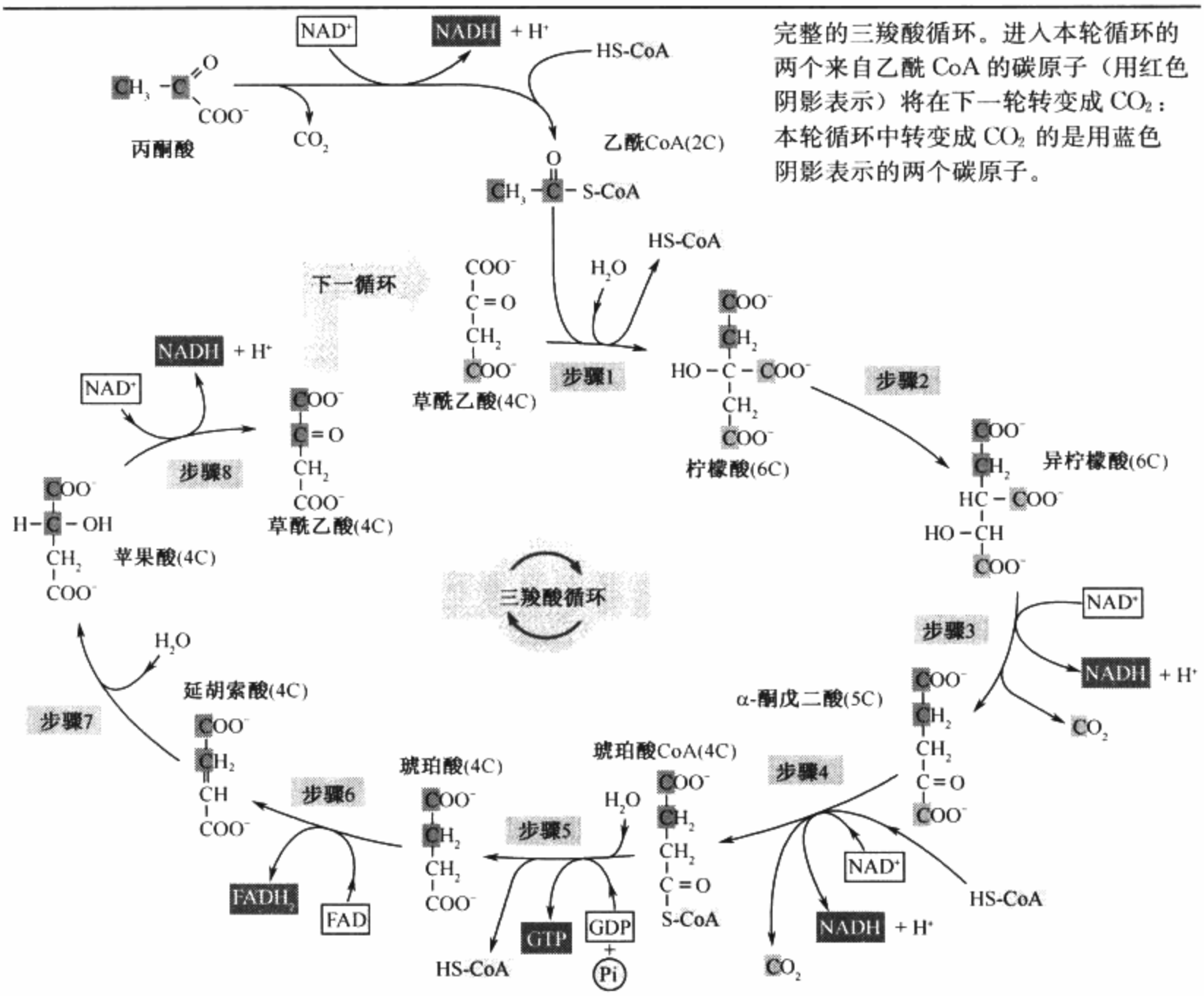
（A）GTP 和 GDP 分别与 ATP 和 ADP 相似。（B）$FADH_2$是氢和高能电子的载体，与 NADH 和 NADPH 类似。这里显示的是它的氧化形式（FAD），用黄色突出能携带氢的原子

程中，电子载体 NADH 和 $FADH_2$将其他分子氧化时，它们所获得的电子转移到包埋在线粒体内膜内的**电子传递链**（electron-transport chain）上。当电子沿这条由特化的电子供体和电子受体组成的电子传递链进行传递时，它们的能量状态连续降低。这个过程中电子释放的能量用于将 H^+（质子）泵过膜，从线粒体内部泵到膜外（图 2-81），因此产生一个 H^+ 梯度。该梯度是能量的来源，像电子一样被用来推动各种需能反应，其中最重要的就是 ADP 磷酸化生成 ATP。

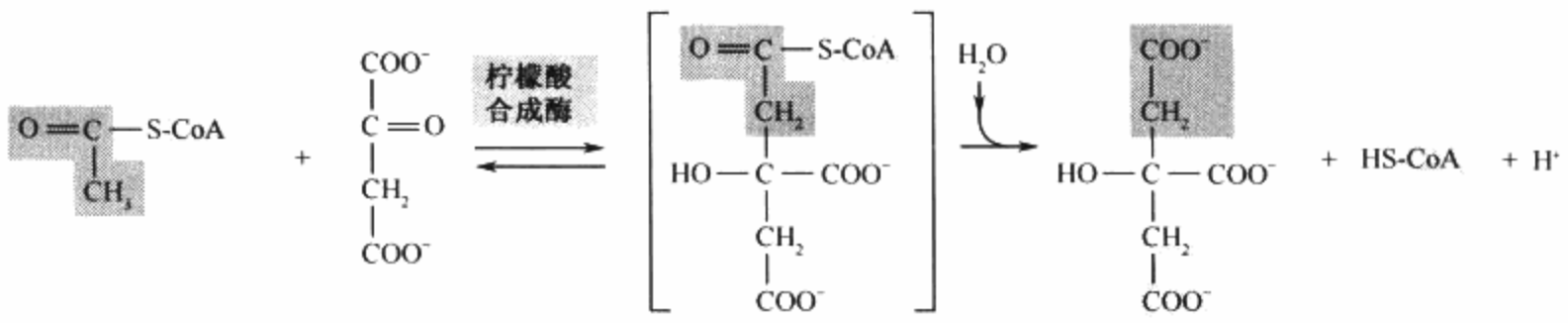
在这一系列电子转移反应最后，电子传递到扩散进入线粒体内的分子氧（O_2），这些分子氧立刻与周围溶液中的 H^+ 结合生成水分子。到此电子处于最低能量状态，因而所有可以获取的能量都从被氧化的食物分子中抽提出来。这个称为**氧化磷酸化**（oxidative phosphorylation）（图 2-82）的过程也在细菌的细胞质膜发生。作为细胞最显著的成功进化之一，第 14 章将主要围绕氧化磷酸化这个主题展开。

总结起来，细胞利用 1 分子葡萄糖完全氧化成 CO_2 和 H_2O，来生成约 30 分子 ATP。相比之下，如果只经过糖酵解途径，每分子葡萄糖仅仅生成 2 分子 ATP。

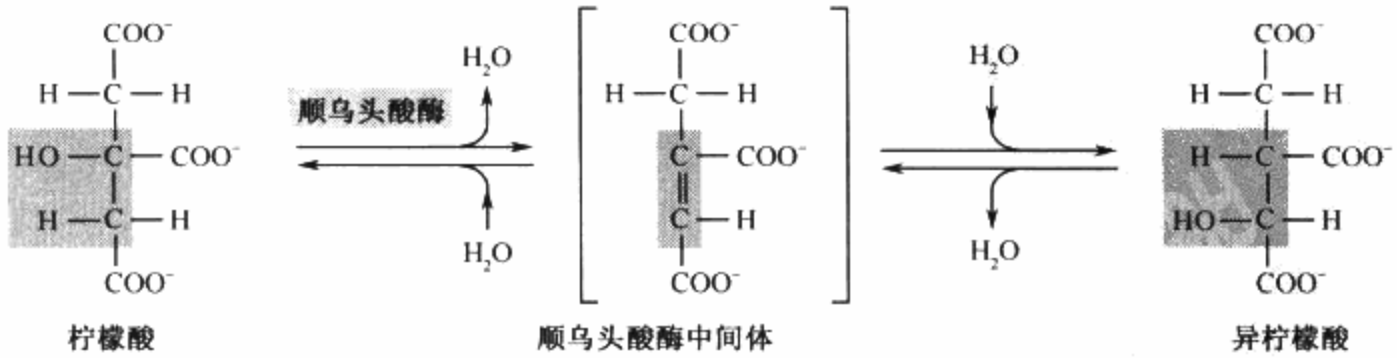
专题 2-9　完整的三羧酸循环



下面是 8 步反应的详细过程。每一步发生变化的分子用蓝色表示，黄色盒中是催化该反应的酶的名称。

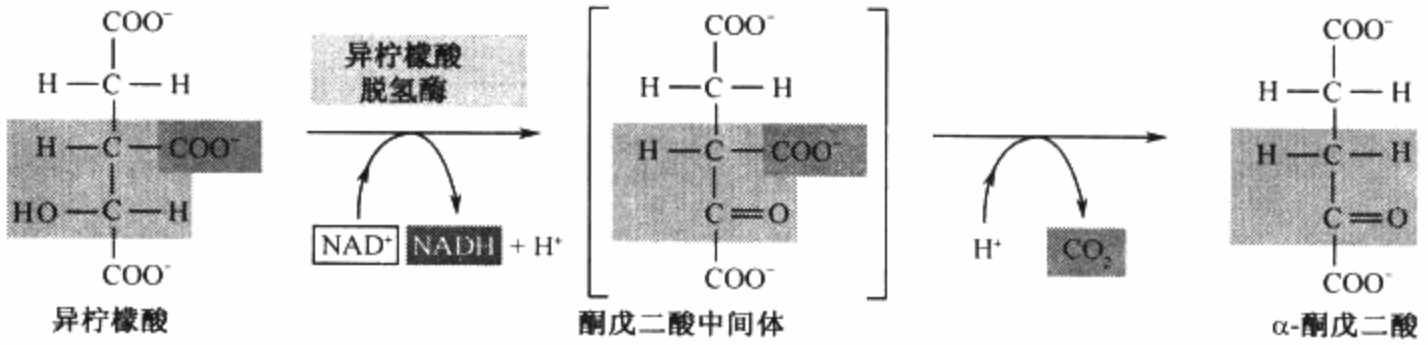


第 1 步：酶从乙酰 CoA 的 CH_3 基团移走一个质子以后，带负电荷的 CH_2^- 与草酰乙酸的羧基碳之间形成共价键。紧接着辅酶 A（CoA）的水解释放的能量推动反应迅速向前。

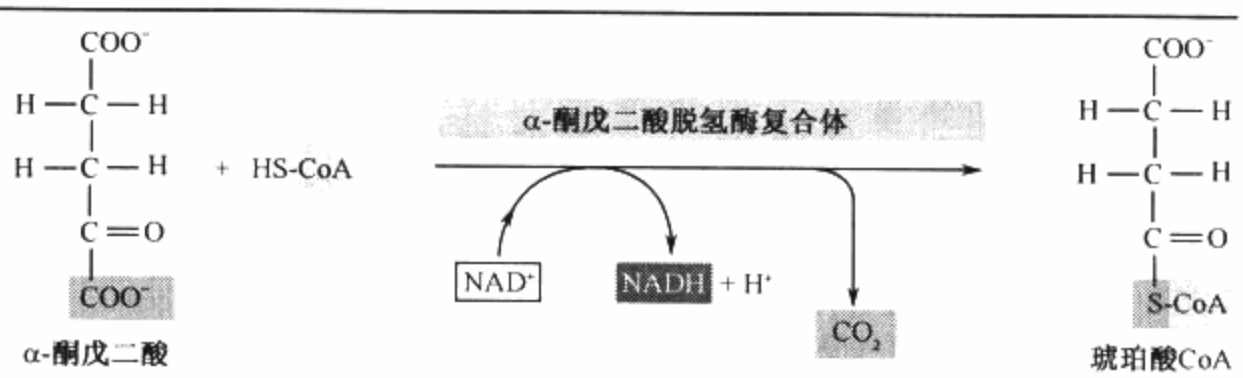


第 2 步：异构化反应过程中先失去水，然后又加上水，结果使羟基从一个碳原子转移到相邻的碳原子上。

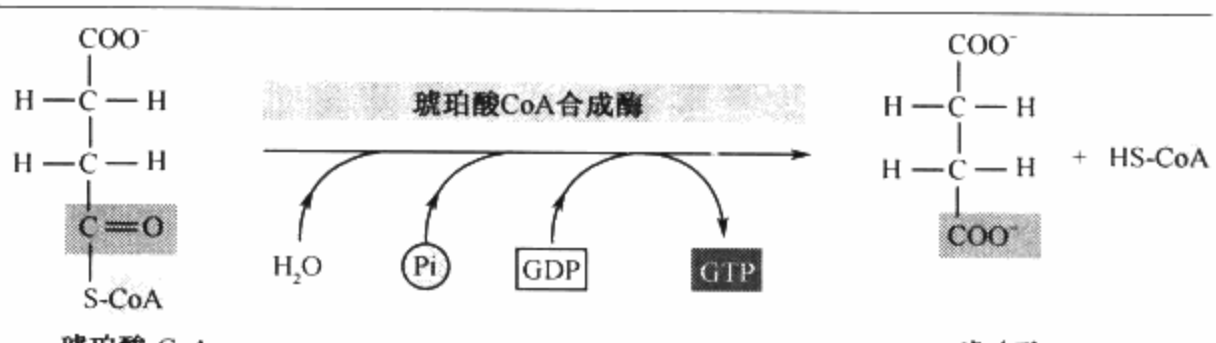
第 3 步：在该循环的 4 步氧化步骤的第一步中，携带羟基的碳转变成羰基。中间产物是不稳定的，当与异柠檬酸脱氢酶结合时失去 CO_2。



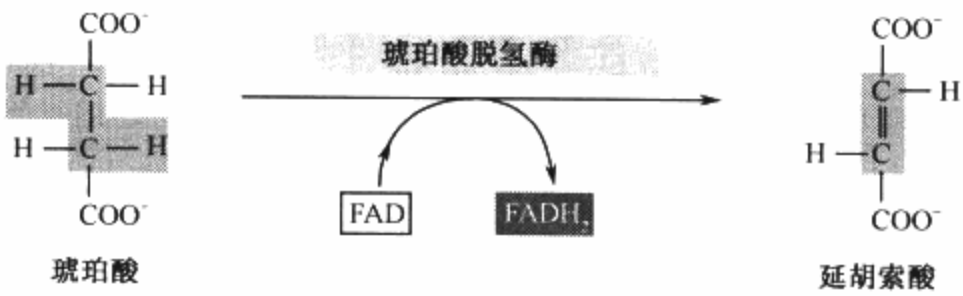
第 4 步：α-酮戊二酸脱氢酶复合体非常像大的酶复合体（丙酮酸脱氢酶），将丙酮酸转变成乙酰 CoA。α-酮戊二酸脱氢酶复合体同样催化一种氧化过程，该氧化过程产生 NADH、CO_2 和与辅酶 A（CoA）连接的高能硫酯键。



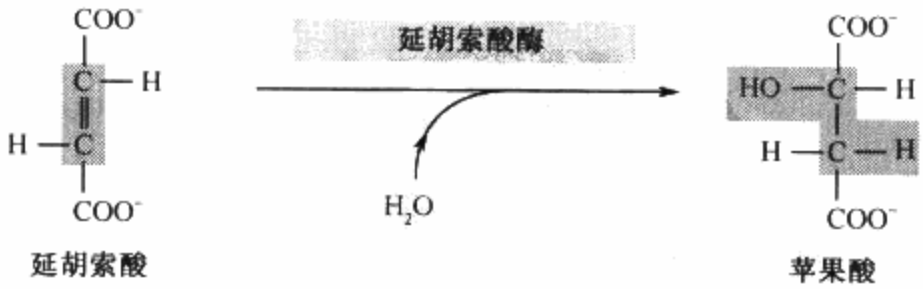
第 5 步：来自溶液的磷酸分子替代 CoA，形成一个与琥珀酸连接的高能磷酸键。然后该磷酸传递给 GDP 以形成 GTP。（在细菌和植物中，形成 ATP。）



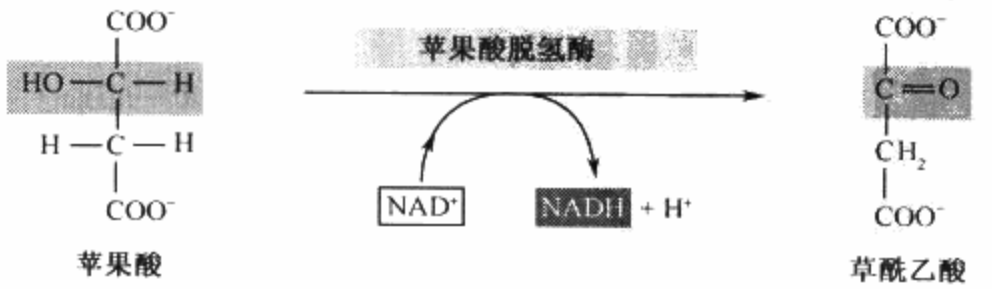
第 6 步：在该循环的第 3 步，FAD 从琥珀酸上移去 2 个氢原子。



第 7 步：延胡索酸的加水使羟基与羰基碳相邻。



第 8 步：在该循环的 4 步氧化步骤的最后一步中，携带羟基的碳转变成羰基，再生成第一步所需的草酰乙酸。



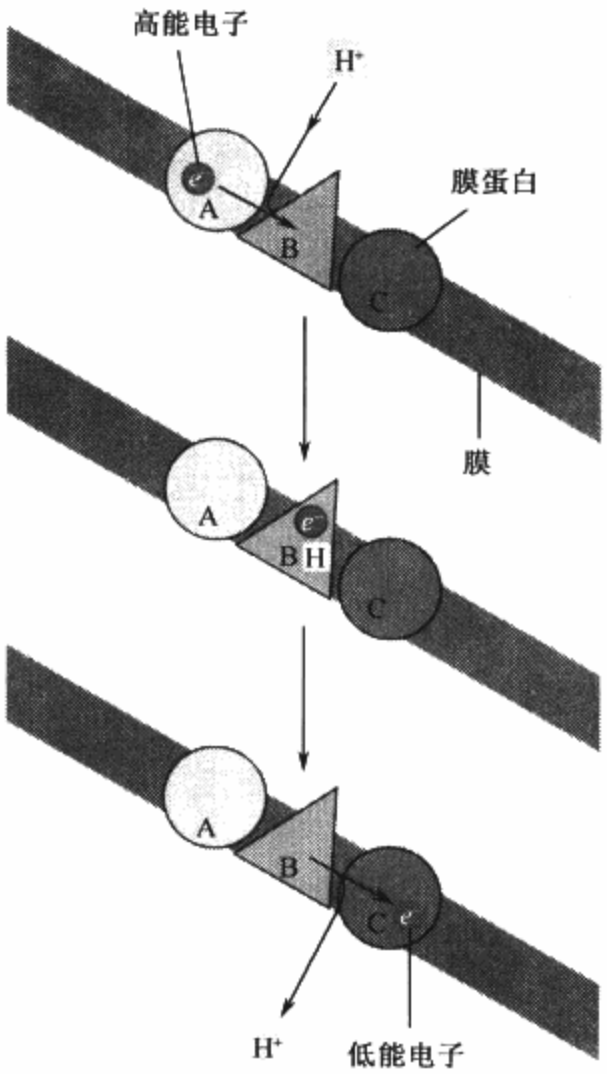


图 2-81　电子转移反应产生的跨膜 H^+ 梯度

高能电子（如来源于某个代谢物的氧化）被载体 A、B 和 C 依次传递到一个较低的能量状态。图中载体 B 位于膜内，当电子经过它时，它能从一边吸收 H^+，而从另一边释放，结果产生 H^+ 梯度。这个梯度代表一种储存的能量，它可以被其他膜蛋白利用，来推动 ATP 的生成

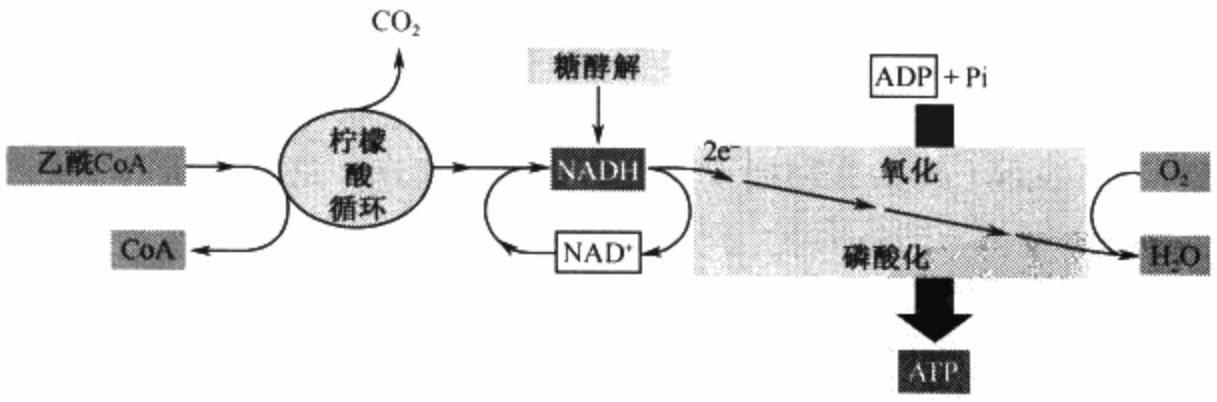


图 2-82　食物分子氧化的最后阶段

NADH 和 $FADH_2$（图中未显示 $FADH_2$）分子均由柠檬酸循环产生。这些活化载体提供高能电子，它们最终用于将 O_2 还原成 H_2O。这些电子沿着线粒体内膜（或者细菌的细胞质膜）电子传递链传递释放的大部分能量被用来推动 ATP 的合成。因此，这个过程称为氧化磷酸化

有机体使用特殊的仓库储存食物分子

所有的生物如果要维持细胞内的生物有序性，就需要维持较高的 ATP/ADP 比例。但是

动物仅仅间歇性地采食，而植物需要度过没有阳光的夜晚，不能通过光合作用合成糖类物质。由于这个原因，植物和动物都将糖和脂肪转化成特殊的形式储存起来（图 2-83）。

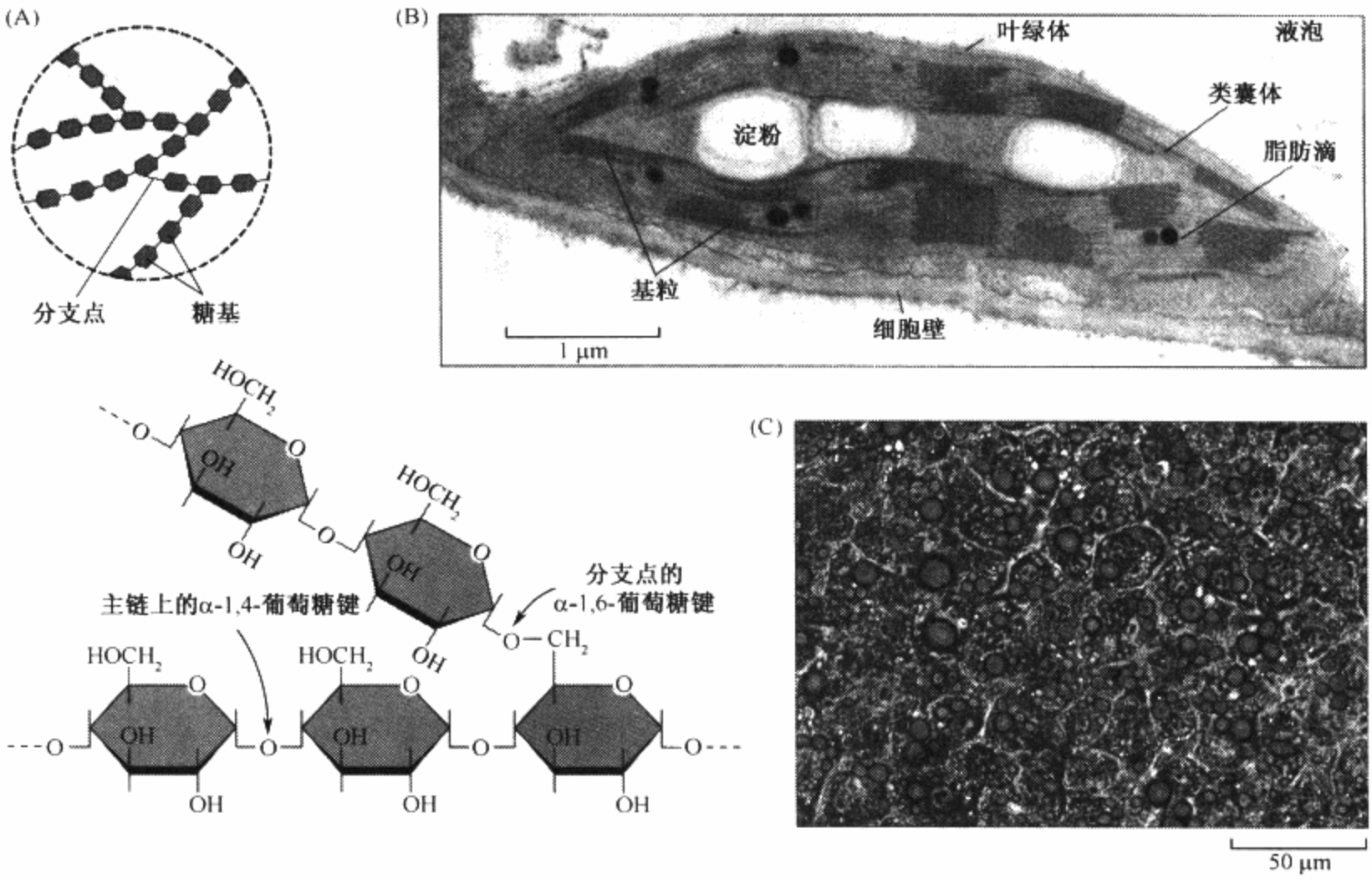


图 2-83　动物和植物细胞内糖和脂肪的储存

（A）淀粉和糖原的结构。淀粉和糖原分别为植物和动物内糖的储存形式。两者都是葡萄糖的储存聚合物，仅仅在分支程度上不同（黄色区域在下面放大显示），糖原比淀粉的分支更多。（B）植物细胞单个叶绿体的一小部分，显示了这里储存了生物合成反应生成的脂肪颗粒和脂肪滴。（C）脂肪滴（涂成红色）在处于发育中的动物细胞内积聚［（B）经 K. Plaskitt 许可；（C）经 Ronald M. Evans 和 Peter Totonoz 许可］

为了抵抗长期挨饿，动物以脂肪滴的形式储存脂肪酸，这些脂肪滴由不溶于水的三酰甘油组成，大部分位于特化的脂肪细胞内。而对于短期储存的糖类物质，则以具有巨大分支的聚糖——**糖原**（glycogen）中的葡萄糖亚单位形式储存起来。这些糖原以小颗粒存在于包括肝和肌肉细胞的许多细胞的细胞质中。根据需要，糖原的合成和降解受到迅速的调控。如果所需的 ATP 比从血液吸收的食物分子产生的 ATP 更多，那么细胞就降解糖原，在此反应中生成葡萄糖-1-磷酸，然后进入糖酵解途径。

从数量上讲，**脂肪**（fat）是比糖原更重要得多的能量储存形成，一部分原因是由于氧化 1g 脂肪释放的能量大约是氧化 1g 糖原的 2 倍。更重要的是，与脂肪不同，糖原结合了大量的水分子，导致同样数量的能量所需糖原与脂肪的数量相差 6 倍。一个普通成年人储存的糖原仅可供一天的活动消耗，而储存的脂肪可供将近一个月。如果我们主燃料库必须储存糖原而不是脂肪，那么我们的体重就将不得不增加大约 60lb（1lb=0.453 592kg）。

我们的脂肪大多数都储存在脂肪组织里，如果需要，就可以从这里释放脂肪到血液中供其他的细胞利用。如果长时间不吃东西就会产生这种需要；即便平时一个晚上不进食也会导致脂肪动员，所以早晨大多数进入柠檬酸循环的乙酰 CoA 大多数来自于食物中的葡萄糖，并且任何过量的葡萄糖都用于补充消耗尽的糖原储存或者是合成脂肪（尽管动物细胞很容易将

糖转变成脂肪，它们却不能将脂肪酸转变成糖）。

植物虽然通过光合作用生成 NAPH 和 ATP，但这个重要的过程只发生在特化的细胞器——叶绿体内。叶绿体被一层对于两种活化载体分子不通透的膜与植物其他细胞隔开。而且植物存在许多缺少叶绿体的细胞，如根中的细胞，因此不能合成自己的糖类和 ATP。因此，对于大多数 ATP 的生成，植物都有赖于从叶绿体向线粒体中糖的运输，后者在所有植物细胞中都存在。大部分植物所需的 ATP 都在线粒体中合成，并且被输送到其他的植物细胞中，其途径与非光合生物中用来氧化降解糖类物质的途径相同（图 2-84）。

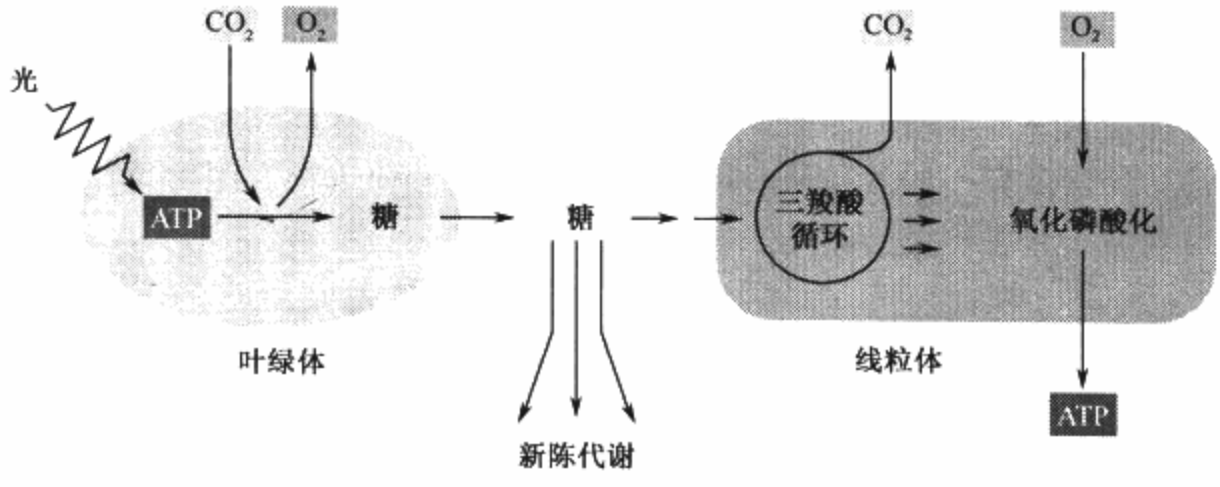


图 2-84　大多数植物细胞代谢所需的 ATP 是如何产生的

在植物中，叶绿体和线粒体协同作用为细胞提供代谢中间物和 ATP

在白天光合作用能力过剩的期间，叶绿体将它们生产的一部分糖转化成脂肪和**淀粉**（starch），淀粉是一类葡萄糖的多聚体，类似于动物中的糖原。植物中的脂肪是三酰甘油，就像动物中的脂肪一样，仅仅是占大多数的脂肪酸种类不同而已。脂肪和淀粉都储存在叶绿体中，作为备用库存以便夜晚所用［图 2-83（B）］。

植物种子里的胚胎依靠储存的能量来存活较长的时期，直到它们发芽并长出叶子，才能利用阳光中的能量。因为如此，植物种子通常含有特别多的脂肪和淀粉，这使得它们成为动物的主要食物来源，包括人类自己（图 2-85）。

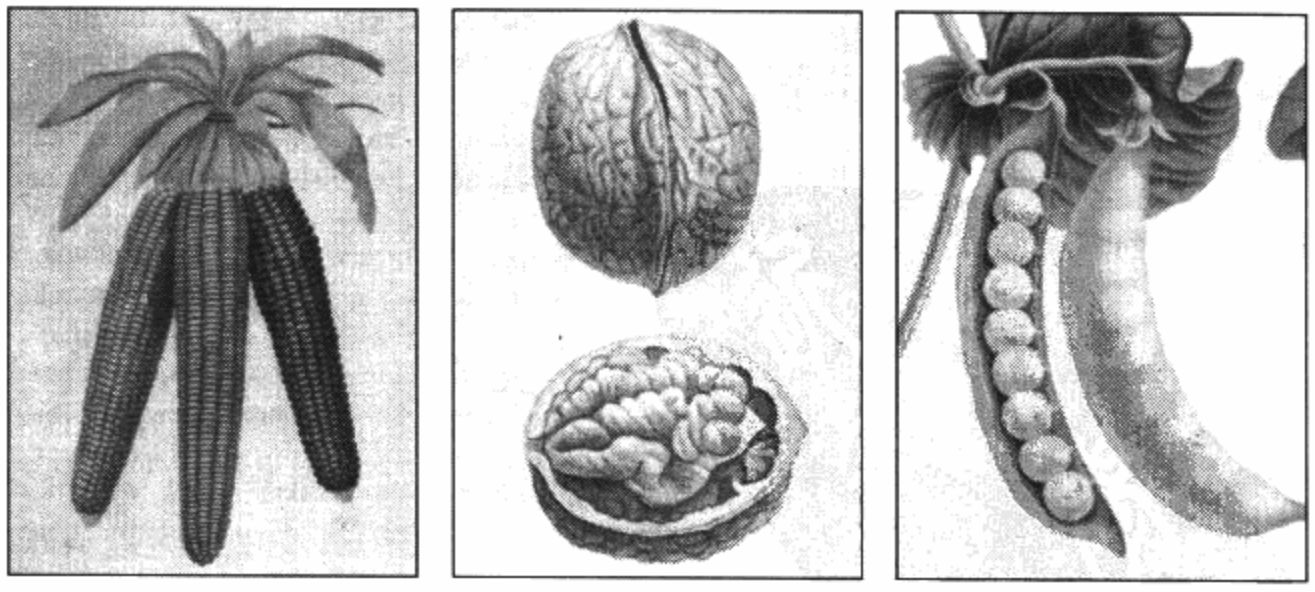


图 2-85　一些作为人类重要食物的植物种子

玉米、核桃仁和豆荚中都含有丰富的淀粉和脂肪，为种子早期的植物胚胎提供能量并构造分子以进行生物合成（经 the John Innes Foundation 允许）

氨基酸和核酸参与了氮循环

到目前为止，我们的讨论都集中在碳水化合物的代谢上。还没有考虑氮或硫的代谢。这两种元素是蛋白质和核酸的组成成分，而蛋白质和核酸则是细胞中两大类最重要的大分子集合，占细胞干重的大约 2/3。通过一系列可逆的循环，氮原子和硫原子从一个复合物转移另一个复合物，也会在生物体和它们的环境之间进行转移。

虽然分子氮在地球大气层中含量丰富，但是氮气在从化学水平上说没有反应活性。仅有一些活生物能够将分子氮合成到有机分子中，这个过程称为**固氮作用**（nitrogen fixation）。固氮作用在某些微生物中存在并通过一些地理物理过程进行，如闪电、放电过程。固氮作用对于整个生物圈是至关重要的。因为如果没有它，生命就不可能在地球上存活。然而，现在生物体中含氮复合物仅有一部分是从一种生物转移到另一种生物。因此可以认为目前的固氮反应对于总的氮供应起着积极的作用。

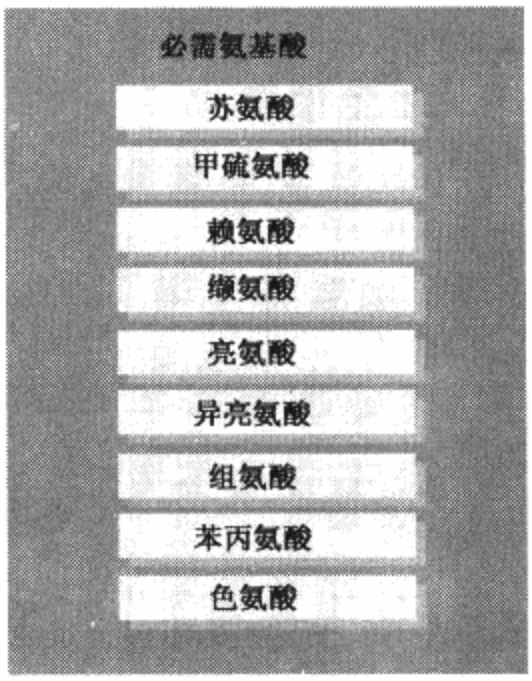


图 2-86　9 种必需氨基酸

人类细胞不能合成这些氨基酸，所以必须由饮食提供

脊椎动物实际上从每天所吃的蛋白质和核苷酸中获取所有的氮。在体内，这些大分子分解成氨基酸和核酸组分，它们所含的氮用来合成新的蛋白质和核酸或者用于合成其他分子。蛋白质中发现的 20 种氨基酸中大约有一半对于脊椎动物来说都是必需氨基酸，也就是说它们不能由膳食中的其他成分合成（图 2-86）。而其他的氨基酸可以这样合成，通过利用各种原材料，包括柠檬酸循环的中间产物，按下面描述的那样进行。必需氨基酸由无脊椎生物合成，通常是经由冗长的能量消费很大的途径进行，这些途径在脊椎动物进化过程中已不存在。

RNA 和 DNA 合成所需的核苷酸可以通过特殊的生物合成途径进行合成：不存在必须由膳食提供的“必需核苷酸”。嘌呤和嘧啶中所有的氮（以及一些碳）都来源于含量丰富的氨基酸，如谷氨酸、天冬氨酸和甘氨酸，而核糖和脱氧核糖则来源于葡萄糖。

生物合成中未利用的氨基酸可以被氧化产生代谢能：它们中的大部分碳和氢原子最终形成 CO_2 和 H_2O，而氮原子经过各种形式的转变，最终出现在尿液中被排出体外。每种氨基酸都由不同的途径进行代谢，这些过程需要一系列酶催反应的存在。

许多生物合成途径起始于糖酵解作用或柠檬酸循环

代谢为细胞提供能量以及合成许多其他分子所需的构建分子（图 2-36）。到现在为止，我们关于糖酵解作用和柠檬酸循环的讨论都强调能量的生成，而未强调它们为生物合成提供起始物质。但是，这些反应途径中形成许多中间产物也被其他的酶催化生成氨基酸、核苷酸、脂质和其他细胞所需的较小的有机分子。图 2-87 列举了从中心代谢反应到生物合成的一些分支途径，从中可看出这个过程极为复杂。

细胞中如此多分支途径的存在要求中间产物在每个分支处的去向受到调节，这些我们将接着讨论。

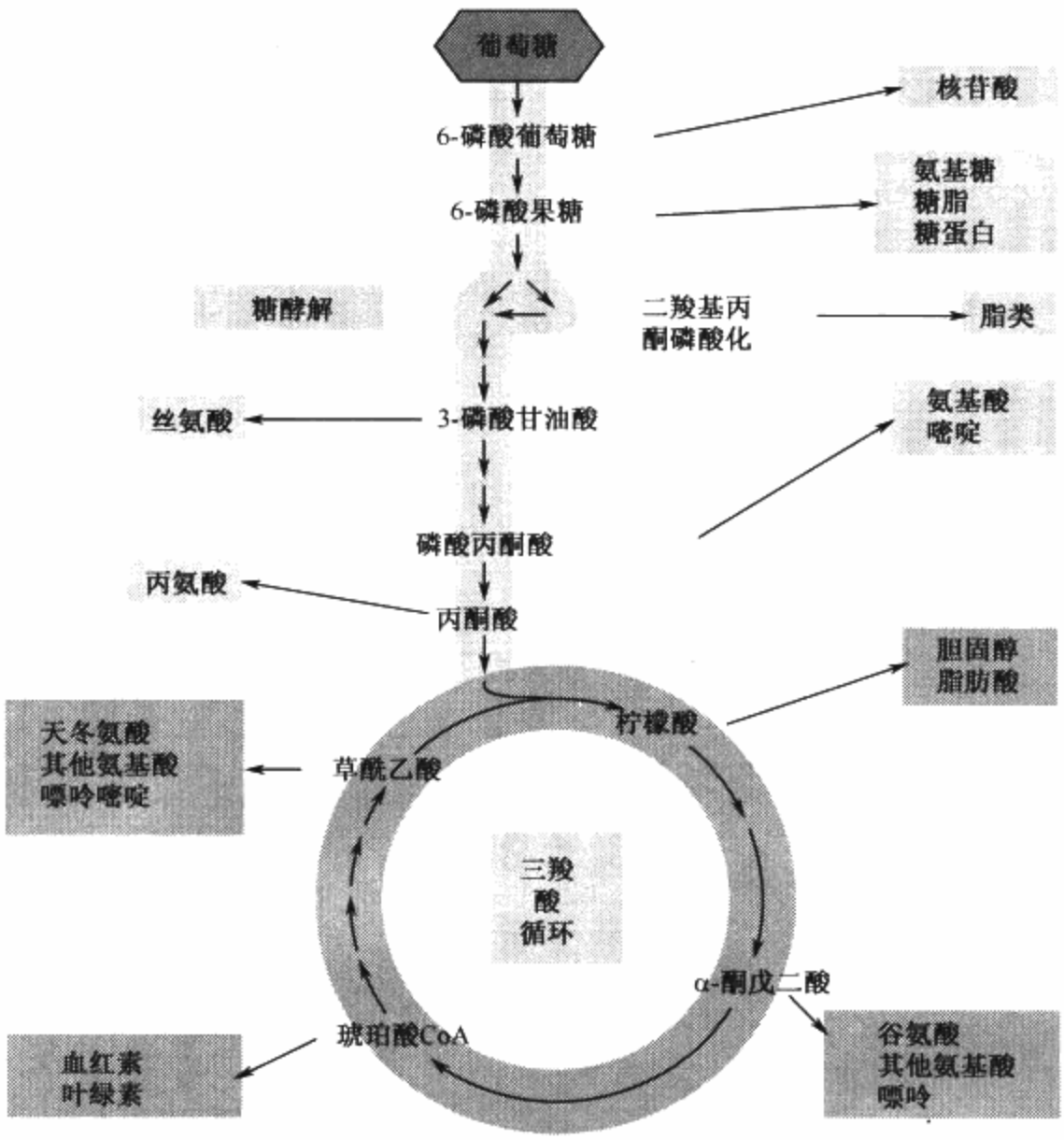


图 2-87　糖酵解途径和柠檬酸循环为许多重要的生物分子合成提供前体

氨基酸、核苷酸、脂类、糖类和其他的分子（这里作为产物列出）反过来作为细胞中许多大分子合成的前体。图中每个黑色箭头表示一个单酶催化的反应；红色箭头通常代表合成所指产物所必需的由许多步骤组成的途径

代谢受到组织和调节

将糖酵解途径和柠檬酸循环与其他代谢途径的关系概括于图 2-88，我们可以感受到细胞作为一个化学机器有多么复杂。这类图表在本章前面用于介绍新陈代谢，仅仅代表细胞中进行的一部分酶促反应，很明显我们对细胞代谢的讨论仅仅涉及细胞化学的一小部分。

所有这些反应都发生在一个直径不到 0.1mm 的细胞中，并且每个反应都需要不同的酶。从图 2-88 可以清楚地看到，同一分子经常可以经由许多不同的反应途径。例如，丙酮酸作为 6 种或者更多酶的底物，受到它们不同方式的修饰作用；第一种酶催化丙酮酸转化为乙酰 CoA，第二种酶催化它转变为草酰乙酸，第三种酶将丙酮酸分子转变为丙氨酸，第四种酶将丙酮酸转变为乳酸等。所有的这些不同途径竞争相同的丙酮酸分子，类似的成千上万其他小分子的竞争同时也在进行着。一张能将代谢途径之间更直接联系起来的三维代谢图可能会让我们对这个复杂过程有更深的认识（图 2-89）。

在多细胞生物中情况就更复杂了。不同类型细胞通常需要某种程度上不同组的酶。不同的组织对于整个生物体的物质组成和化学性质也发挥着明显不同的作用。除了特殊的产物如激素或者抗体存在差别之外，在同一生物的不同细胞中“共同的”代谢途径之间也存在显著差异。

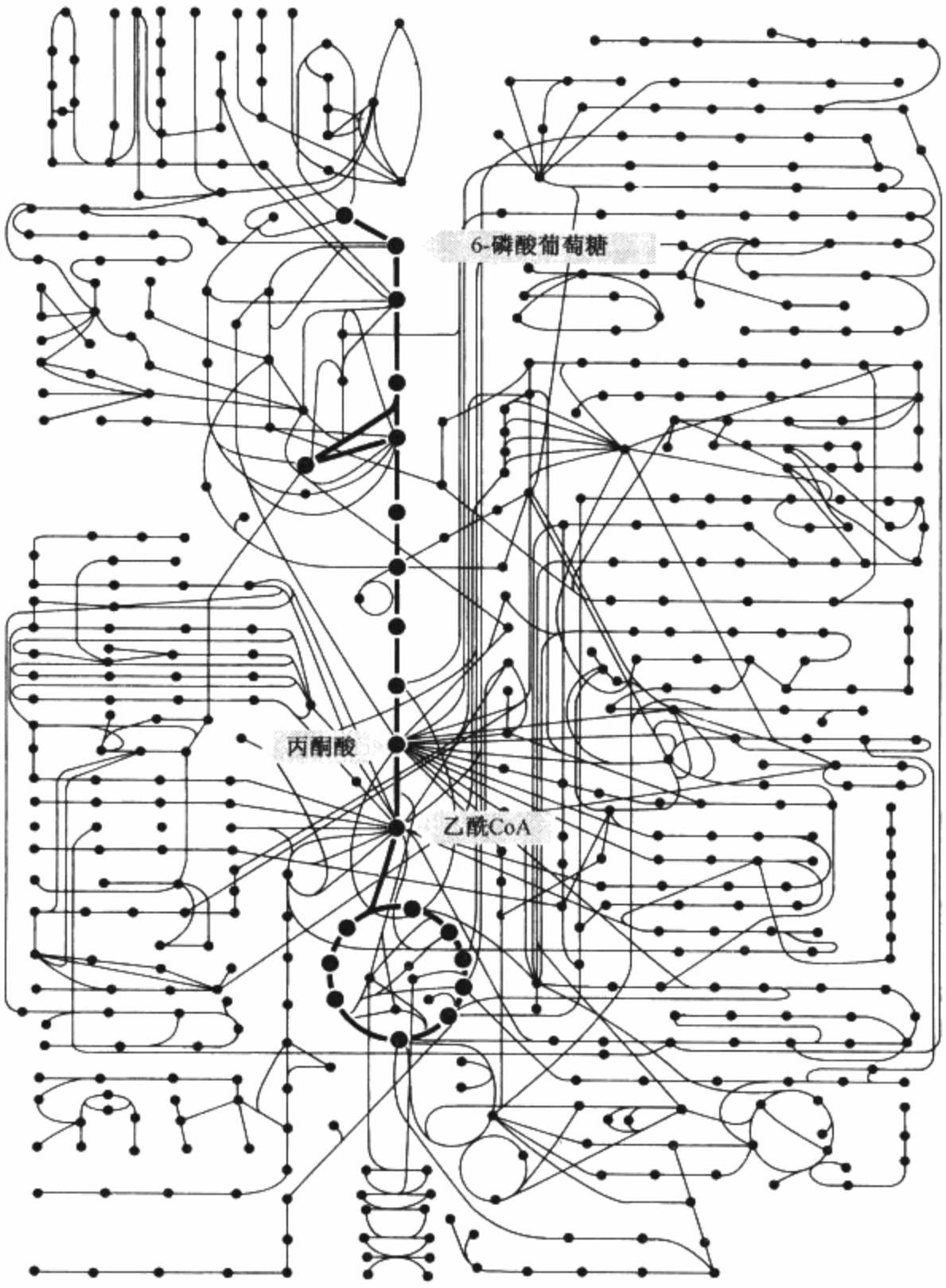


图 2-88　糖酵解和柠檬酸循环处于新陈代谢中的地位

典型细胞中大约 500 种代谢反应用线条表示出来，红色表示糖酵解作用和柠檬酸循环。其他反应或者进入这两个中心途径，释放小分子进行代谢，产生能量；或者离开中心途径因而为生物合成提供碳复合物

尽管实际上所有细胞都含有糖酵解作用、柠檬酸循环、脂类合成与降解、氨基酸代谢所需的酶，但不同组织中对这些过程需求水平却不相同。例如，神经细胞作为体内可能最疲劳的细胞，它没有糖原或脂肪酸储备，几乎完全依赖血流迅速提供葡萄糖。相比之下，肝脏细胞提供葡萄糖给活跃收缩的肌肉细胞，再将肌肉细胞产生的乳酸转变为葡萄糖（图 2-90）。所有类型的细胞都有它们自身显著的代谢特征，在正常状态以及受压力刺激，如饥饿时，它们都广泛地协同作用。也许我们会想到：整个系统必须得到如此精确的平衡，否则的话任何一点微小的错乱，如饮食摄取的暂时改变，都将成为灾难。

事实上，细胞代谢的平衡令人难以置信地稳定。无论什么时候这种平衡受到扰乱，细胞都做出反应以重新恢复到最初的状态。在疾病或饥饿状态时，细胞可以调节自身以适应并继续发

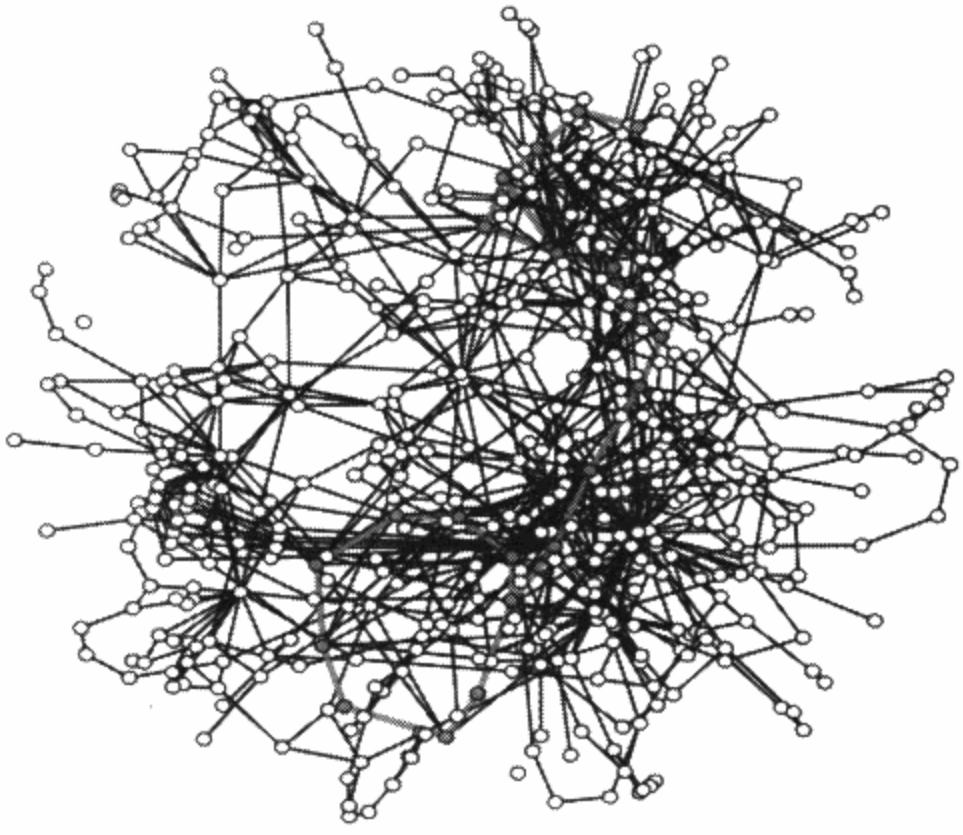


图 2-89　酵母细胞中所有已知涉及小分子的代谢反应

与图 2-88 一样，用红色表明糖酵解反应和柠檬酸循环。这张代谢图的极不寻常之处在于使用了三维结构，所以能够强调许多反应途径之间的相互影响。该图需要在电脑屏幕上观看，在屏幕上它可以旋转，并可以从任何角度进行观察（摘自 H. Jeong, S. P. Mason, A-L. Barabási and N. Oltava, *Nature* 411：41～42，2001. © Macmillan Magazines Ltd.）

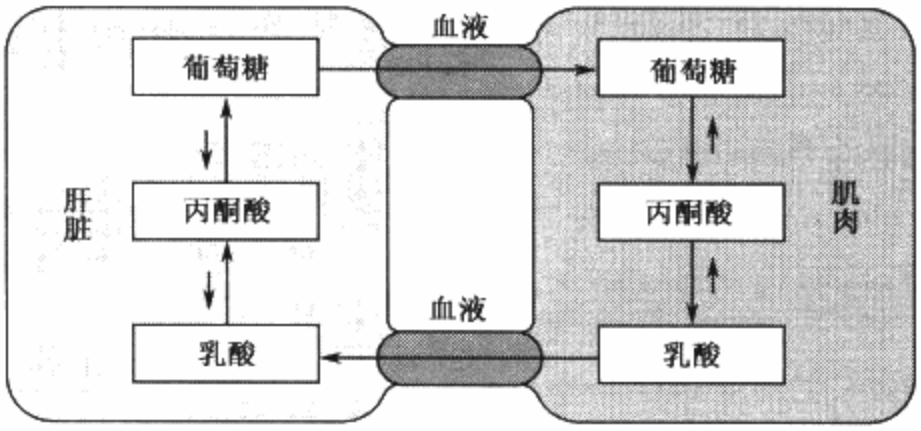


图 2-90　肝脏和肌肉细胞之间代谢协同作用的概览

活跃收缩的肌肉细胞的主要燃料是葡萄糖，它们大多数由肝脏提供。乳酸是肌肉中糖酵解作用使葡萄糖无氧分解的终产物，乳酸在肝脏中通过糖原发生作用转变成葡萄糖

挥作用。许多种突变能破坏甚至完全消除某种反应途径，然而，假如某种最低的要求得到满足，细胞就存活了。细胞能存活是因为一个精细的控制机制网络调节并且协同了细胞内所有反应的速率。这些控制最终有赖于蛋白质根据周围环境的变化改变自身形状和化学特性的显著能力。有关大分子，例如，蛋白质如何合成的原理以及它们受调控的化学机制，我们将在下一章讨论。

小结

葡萄糖和其他食物分子通过有控制的一步一步的氧化作用进行降解，以 ATP 和 NADH 的形式提供化学能。糖酵解作用（在细胞质发生）、柠檬酸循环（在线粒体基质发生）和氧化磷酸化（在线粒体内膜发生）是按顺序发生的三个主要的反应阶段。糖酵解作用和柠檬酸循

环中的中间产物既用作代谢能量的来源，也用来生成许多小分子，作为生物合成的原料。动物细胞以糖原的形式储存糖分子，植物细胞则以淀粉的形式储存糖分子。无论动物和植物都广泛以脂肪作为食物储备。反过来这些储备物质，与占我们吃进去的细胞干物质大多数的蛋白质一起，作为人类主要的食物来源。

参考文献

一般文献

Garrett RH & Grisham CM (1998) Biochemistry, 2nd edn. Orlando: Saunders.

Horton HR, Moran LA, Ochs RS et al. (2001) Principles of Biochemistry, 3rd edn. Upper Saddle River, NJ: Prentice Hall.

Lehninger AL, Nelson DL & Cox MM (1993) Principles of Biochemistry, 2nd edn. New York: Worth.

Mathews CK, van Holde KE & Ahern K-G (2000) Biochemistry, 3rd edn. San Francisco: Benjamin Cummings.

Moore JA (1993) Science As a Way of Knowing. Cambridge, MA: Harvard University Press

Stryer L (1995) Biochemistry, 4th edn. New York: WH Freeman.

Voet D, Voet JG & Pratt CW (1999) Fundamentals of Biochemistry. New York: Wiley.

Zubay GL (1998) Biochemistry, 4th edn. Dubuque, IO: William C Brown.

The Chemical Components of a Cell

Abeles RH, Frey PA & Jencks WP (1992) Biochemistry. Boston: Jones & Bartlett.

Atkins PW (1996) Molecules. New York: WH Freeman.

Branden C & Tooze J (1999) Introduction to Protein Structure, 2nd edn. New York: Garland Publishing.

Bretscher MS (1985) The molecules of the cell membrane. *Sci. Am.* 253(4), 100–109.

Burley SK & Petsko GA (1988) Weakly polar interactions in proteins. *Adv. Protein Chem.* 39, 125–189.

Eisenberg D & Kauzman W (1969) The Structure and Properties of Water. Oxford: Oxford University Press.

Fersht AR (1987) The hydrogen bond in molecular recognition. *Trends Biochem. Sci.* 12, 301–304.

Franks F (1993) Water. Cambridge: Royal Society of Chemistry.

Henderson LJ (1927) The Fitness of the Environment, 1958 edn. Boston: Beacon.

Ingraham JL, Maaløe O & Neidhardt FC (1983) Growth of the Bacterial Cell. Sunderland, MA: Sinauer.

Pauling L (1960) The Nature of the Chemical Bond, 3rd edn. Ithaca, NY: Cornell University Press.

Saenger W (1984) Principles of Nucleic Acid Structure. New York: Springer.

Sharon N (1980) Carbohydrates. *Sci. Am.* 243(5), 90–116.

Stillinger FH (1980) Water revisited. *Science* 209, 451–457.

Tanford C (1978) The hydrophobic effect and the organization of living matter. *Science* 200, 1012–1018.

Tanford C (1980) The Hydrophobic Effect. Formation of Micelles and Biological Membranes, 2nd edn. New York: John Wiley.

Catalysis and the Use of Energy by Cells

Atkins PW (1994) The Second Law: Energy, Chaos and Form. New York: Scientific American Books.

Atkins PW (1998) Physical Chemistry, 6th edn. Oxford: Oxford University Press.

Berg HC (1983) Random Walks in Biology. Princeton, NJ: Princeton University Press.

Dickerson RE (1969) Molecular Thermodynamics. Menlo Park, CA: Benjamin Cummings.

Dressler D & Potter H (1991) Discovering Enzymes. New York: Scientific American Library.

Einstein A (1956) Investigations on the Theory of Brownian Movement. New York: Dover.

Eisenberg D & Crothers DM (1979) Physical Chemistry with Applications to the Life Sciences. Menlo Park, CA: Benjamin Cummings.

Fruton JS (1999) Proteins, Enzymes, Genes: The Interplay of Chemistry and Biology. New Haven: Yale University Press.

Goodsell DS (1991) Inside a living cell. *Trends Biochem. Sci.* 16, 203–206.

Karplus M & McCammon JA (1986) The dynamics of proteins. *Sci. Am.* 254(4), 42–51.

Karplus M & Petsko GA (1990) Molecular dynamics simulations in biology. *Nature* 347, 631–639.

Kauzmann W (1967) Thermodynamics and Statistics: with Applications to Gases. In Thermal Properties of Matter, vol 2, New York: Benjamin.

Klotz IM (1967) Energy Changes in Biochemical Reactions. New York: Academic Press.

Kornberg A (1989) For the Love of Enzymes. Cambridge, MA: Harvard University Press.

Lavenda BH (1985) Brownian Motion. *Sci. Am.* 252(2), 70–85.

Lawlor DW (2001) Photosynthesis, 3rd edn. Oxford: BIOS.

Lehninger AL (1971) The Molecular Basis of Biological Energy Transformations, 2nd edn. Menlo Park, CA: Benjamin Cummings.

Lipmann F (1941) Metabolic generation and utilization of phosphate bond energy. *Adv. Enzymol.* 1, 99–162.

Lipmann F (1971) Wanderings of a Biochemist. New York: Wiley.

Nisbet EE & Sleep NH (2001) The habitat and nature of early life. *Nature* 409, 1083–1091.

Racker E (1980) From Pasteur to Mitchell: a hundred years of bioenergetics. *Fed. Proc.* 39, 210–215.

Schrödinger E (1944 & 1958) What is Life?: The Physical Aspect of the Living Cell and Mind and Matter, 1992 combined edn. Cambridge: Cambridge University Press.

van Holde KE, Johnson WC & Ho PS (1998) Principles of Physical Biochemistry. Upper Saddle River, NJ: Prentice Hall.

Walsh C (2001) Enabling the chemistry of life. *Nature* 409, 226–231.

Westheimer FH (1987) Why nature chose phosphates. *Science* 235, 1173–1178.

Youvan DC & Marrs BL (1987) Molecular mechanisms of photosynthesis. *Sci. Am.* 256(6), 42–49.

How Cells Obtain Energy from Food

Cramer WA & Knaff DB (1990) Energy Transduction in Biological Membranes. New York: Springer-Verlag.

Fell D (1997) Understanding the Control of Metabolism. London: Portland Press.

Flatt JP (1995) Use and storage of carbohydrate and fat. *Am. J. Clin. Nutr.* 61, 952S–959S.

Fothergill-Gilmore LA (1986) The evolution of the glycolytic pathway. *Trends Biochem. Sci.* 11, 47–51.

Huynen MA, Dandekar T & Bork P (1999) Variation and evolution of the citric-acid cycle: a genomic perspective. *Trends Microbiol.* 7, 281–291.

Kornberg HL (1987) Tricarboxylic acid cycles. *Bioessays* 7, 236–238.

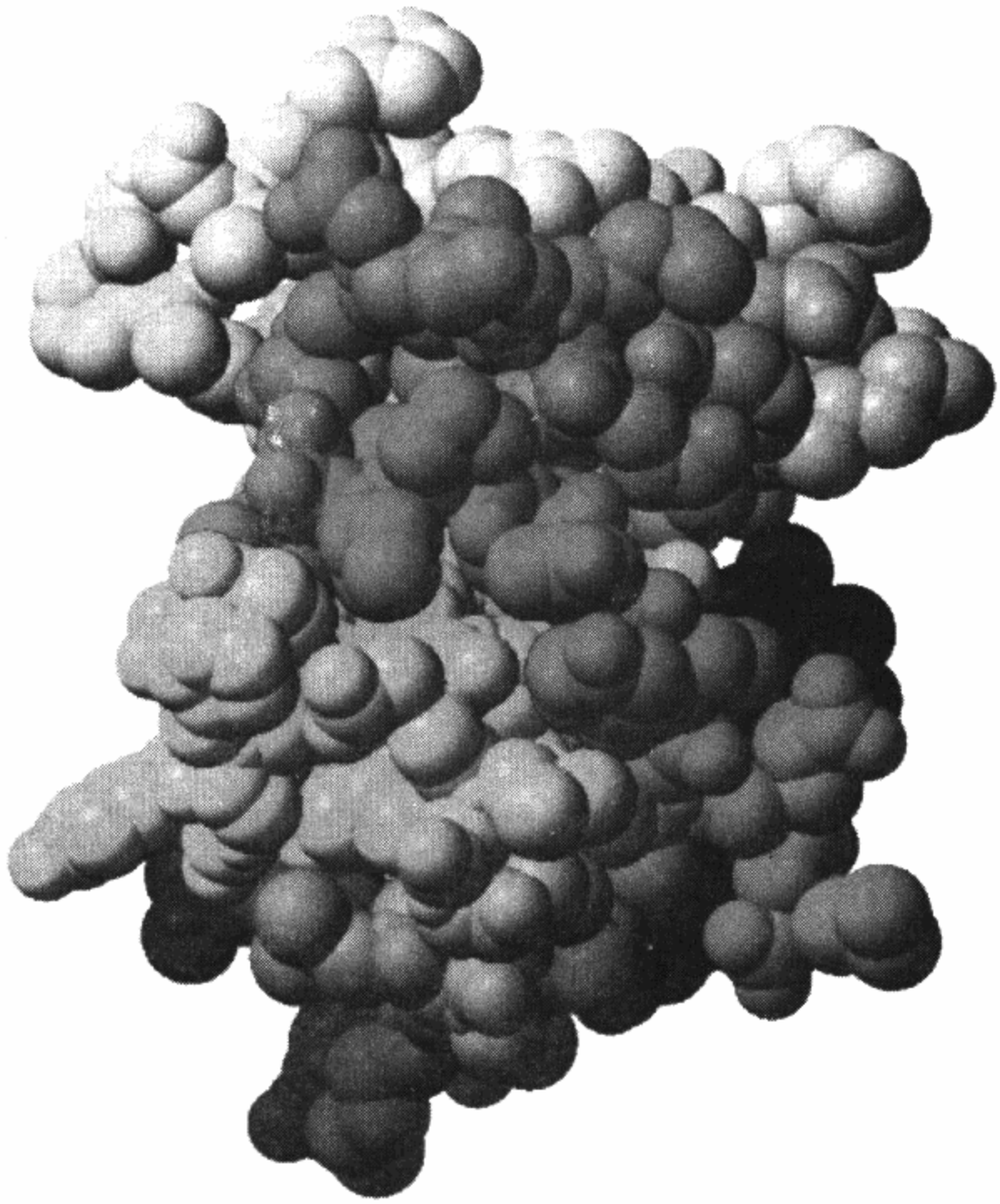
Krebs HA & Martin A (1981) Reminiscences and Reflections. Oxford/New York: Clarendon Press/Oxford University Press.

Krebs HA (1970) The history of the tricarboxylic acid cycle. *Perspect. Biol. Med.* 14, 154–170.

Martin BR (1987) Metabolic Regulation: A Molecular Approach. Oxford: Blackwell Scientific.

McGilvery RW (1983) Biochemistry: A Functional Approach, 3rd edn. Philadelphia: Saunders.

Newsholme EA & Stark C (1973) Regulation in Metabolism. New York: Wiley.



蛋白质结构。图中显示了一个小的蛋白结构域，SH_2 结构域的空间填充图，更多信息请见专题 3-2.（经 David Lawson 许可）

3 蛋 白 质

蛋白质的形状和结构
蛋白质的功能

当我们通过显微镜来观察一个细胞或者分析其物理化学活性时，从本质上来说，其实我们是在观察蛋白质。蛋白质构成了一个细胞干重的绝大部分。它们不仅构成了细胞骨架，而且执行了细胞几乎所有的功能。细胞中的酶提供了复杂的分子表面来帮助其进行很多化学反应。嵌入质膜的蛋白质会形成通道，控制小分子进出细胞。一些蛋白质携带信息从一个细胞传递到另一个细胞，或者作为信号调节者将细胞质膜内的系列信号传递到细胞核。其余一些蛋白质作为小的运动分子机器，如动力蛋白（kinesin），驱动细胞器通过细胞质；拓扑异构酶（topoisomerase）可以解开缠绕在一起的 DNA 分子。其余一些特殊的蛋白质可以作为抗体、毒素、激素、防冻剂分子、弹性纤维、绳索以及发光原料。在我们希望理解基因是怎么工作、肌肉如何收缩、神经如何导电、胚胎如何发育，或者我们的身体是怎么发挥功能之前，我们必须对蛋白质有一个深入的认识。

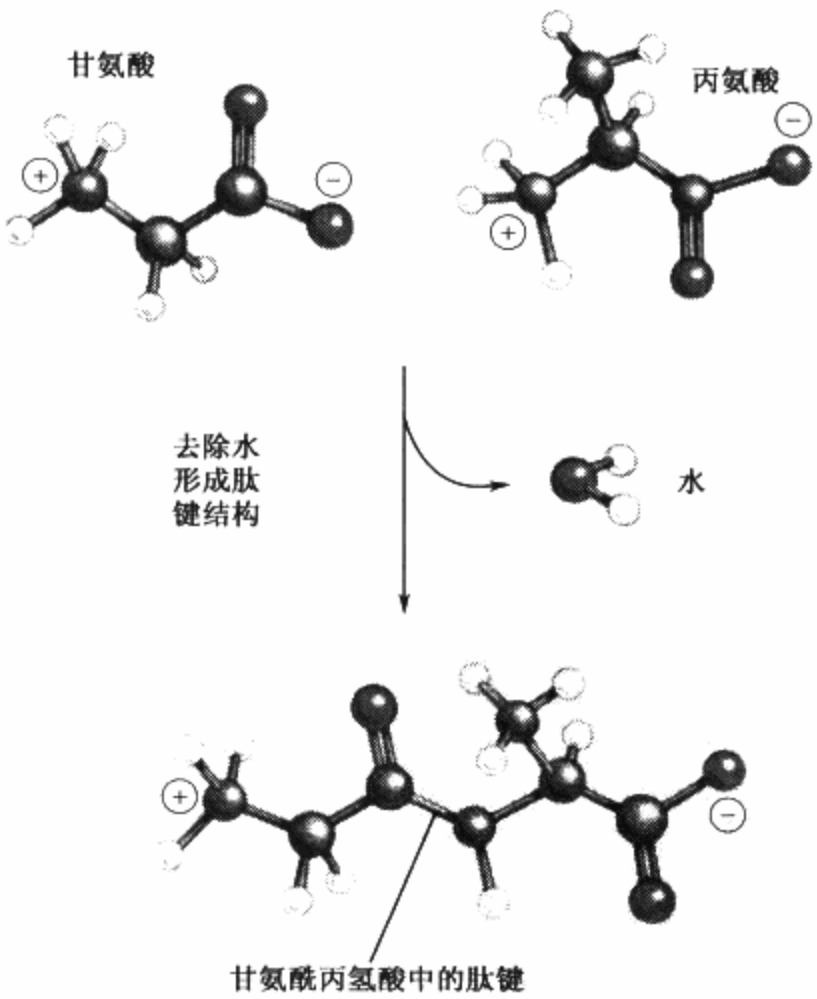


图 3-1　肽键

一个氨基酸的羧基碳原子与第二个氨基酸的氨基氮原子（蓝色）共享电子时形成肽键。如图所示，在这个缩合反应中会失去 1 分子的水

蛋白质的形状和结构

从化学的角度来看，蛋白质是迄今为止发现的结构最复杂和功能最微妙的分子。但当我们意识到每种蛋白质的结构和化学性质在上亿年的进化史中不断发展和调整时，我们会觉得这可能并不稀奇。在这一章开始会考虑形成蛋白质的一长串氨基酸链上的每一个氨基酸的定位是如何决定蛋白质的三维形状。接着我们会利用从原子水平上理解蛋白质的结构来解释蛋白质分子的精确结构如何决定其在细胞中的功能。

蛋白质的形状特异性决定于其氨基酸序列

回顾第 2 章，组成蛋白质的主要是 20 种氨基酸，每一种氨基酸具有不同的化学性质。一个蛋白质分子由这些氨基酸形成长链构成，每一个氨基酸通过共价肽键与其相邻的分子连接（图 3-1）。

因此蛋白质也被叫做多肽。每一种蛋白质有一个独特的氨基酸序列，同种蛋白质中的每个分子是完全相同的。在已知的成千上万种蛋白质中，每一种都有独特的氨基酸序列。

核心多肽链上的重复原子序列称作**多肽链骨架**。与多肽链骨架相连的氨基酸的另一部分没有参与形成肽键，它们赋予了每种氨基酸独特的性质：20 种不同的**氨基酸侧链**（图 3-2）。有一些侧链是非极性和疏水的，有一些带有正电或负电，有一些具有反应性等。它们的原子结构在专题 3-1 中显示，在图 3-3 中有一个缩写的简单列表。

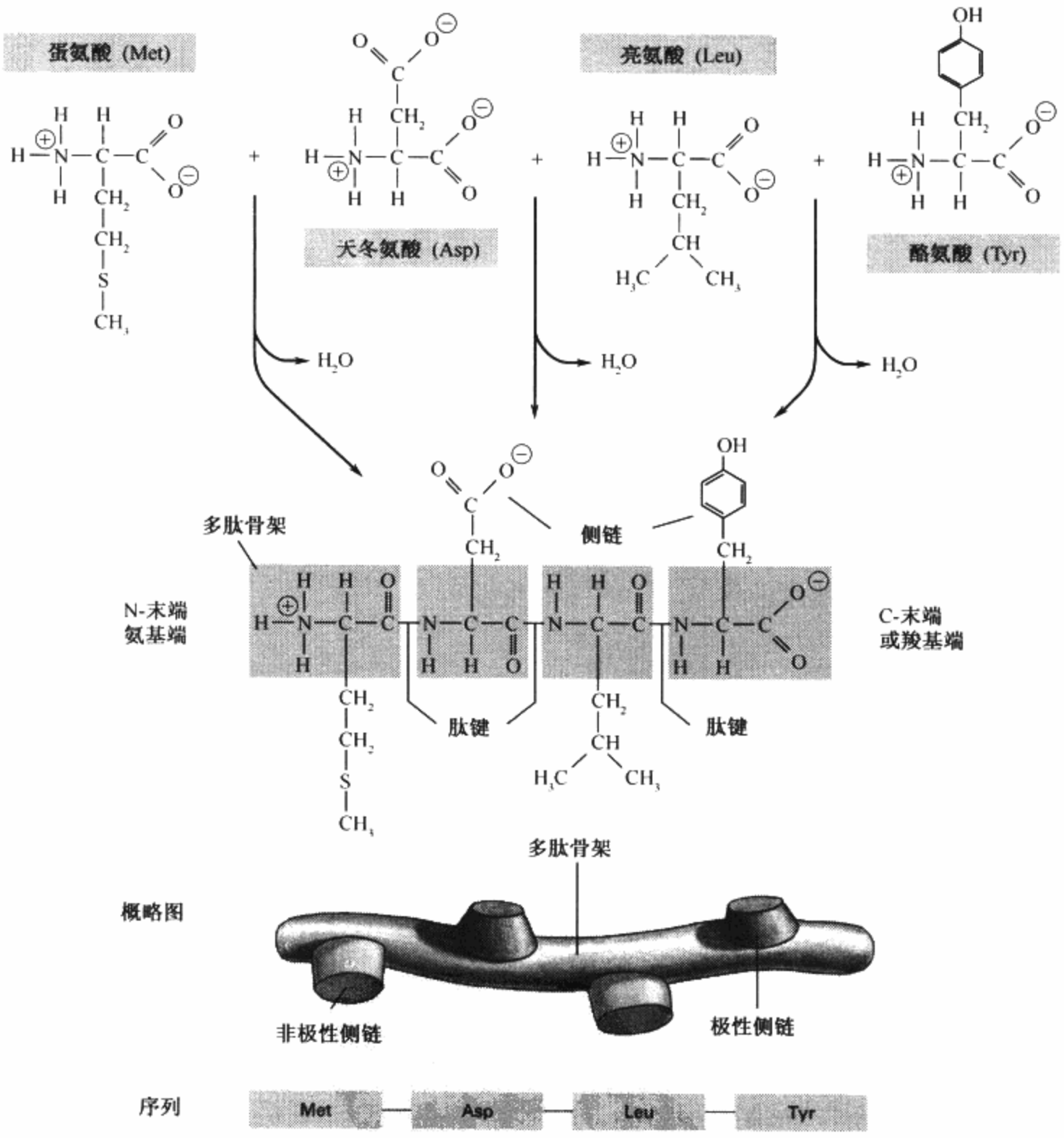


图 3-2　一个蛋白质的结构组成

一个蛋白质由一条多肽链骨架和结合在其上的侧链组成。每一种蛋白质在氨基酸序列和数量上不相同；因此，化学性质不同的侧链的序列决定了蛋白质的特异性。多肽链的两个末端是不相同的：有自由氨基（NH_3^+）的末端是氨基端或 N 端，有自由羧基的末端是羧基端或 C 端。蛋白质的氨基酸序列通常是 N 到 C 的方向，从左往右读

正如在第 2 章所讨论的那样，原子的行为就像它们是一个有着确定半径（范德华半径）的坚硬球体。任何两个原子不相互重叠的要求极大的限制了一条多肽链可能的键角（图 3-4）。

这个限制以及其他一些空间的相互作用严重地限制了原子的三维空间排列［构象（conformation）］的多样性。不过，一条长的松散的链，如蛋白质，仍然可以有多种折叠方式。

氨基酸			侧链
天冬氨酸	Asp	D	负电
谷氨酸	Glu	E	负电
精氨酸	Arg	R	正电
赖氨酸	Lys	K	正电
组氨酸	His	H	正电
天冬酰胺	Asn	N	不带电极性
谷氨酰胺	Gln	Q	不带电极性
丝氨酸	Ser	S	不带电极性
苏氨酸	Thr	T	不带电极性
酪氨酸	Tyr	Y	不带电极性

极性氨基酸

氨基酸			侧链
丙氨酸	Ala	A	非极性
甘氨酸	Gly	G	非极性
缬氨酸	Val	V	非极性
亮氨酸	Leu	L	非极性
异亮氨酸	Ile	I	非极性
脯氨酸	Pro	P	非极性
苯丙氨酸	Phe	F	非极性
甲硫氨酸	Met	M	非极性
色氨酸	Trp	W	非极性
半胱氨酸	Cys	C	非极性

非极性氨基酸

图 3-3　组成蛋白质的 20 种常见氨基酸

三字母和单字母简写都列出。如图所示极性和非极性侧链的数目相同。原子结构图见专题 3-1

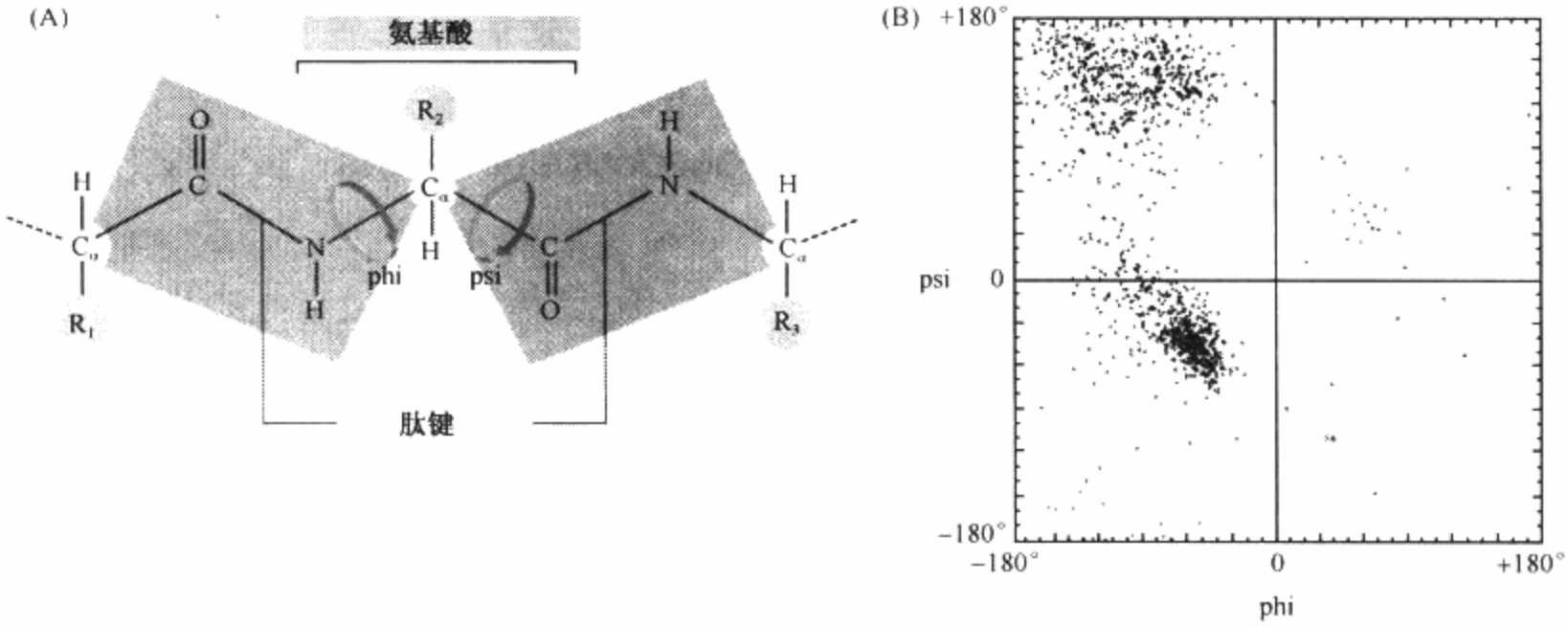


图 3-4　多肽链中空间排列对键角的限制

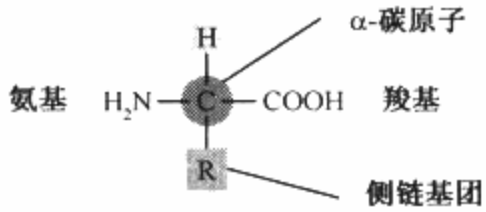
（A）每个氨基酸对多肽链骨架贡献三个键（红色）。肽键是平面的（暗灰色）而且不允许旋转。相反地，C_α—C 是可以旋转的，旋转角度叫做 psi（ψ），N—C_α 的旋转角度叫做 phi（ϕ）。按习惯，R 通常代表氨基酸的侧链基团（绿环）。（B）蛋白质中主链上原子的构象取决于每个氨基酸的 ϕ 角和 ψ 角；因为每个氨基酸内原子的空间冲突，许多 ϕ 和 ψ 角不能形成。在这个称为 R 的图中，每个点代表了一个蛋白质中所观察到的一对角

蛋白质链的折叠也受到存在于链内的多种弱的非共价键的限制。它涉及多肽链骨架和侧链上的原子。这些弱的键包括三种：氢键、离子键和范德华力（这些在第 2 章中有解释）。每个非共价键比典型的构成生物分子的共价键弱 30～300 倍。但是大量弱的作用力的平行作用可以使一条多肽链的两个区域紧密地结合在一起。因此每种折叠方式的稳定性取决于大量这种非共价键的群体效应（图 3-5）。

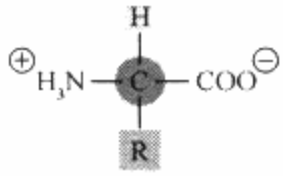
专题 3-1　蛋白质中的 20 种氨基酸

氨基酸

氨基酸的通式如下

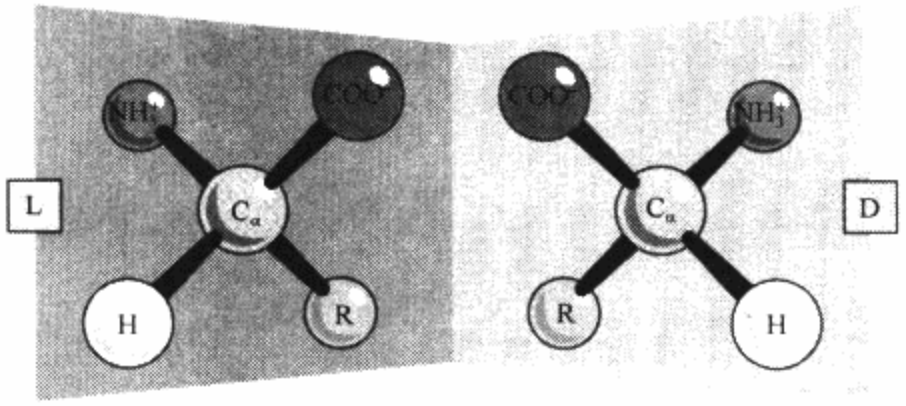


R 通常是 20 种不同侧链中的一种，pH 为 7 时氨基和羧基离子化



光学异构体

α 碳原子是非对称的，因此有两个镜像异构体，L 和 D。



蛋白专有 L-氨基酸

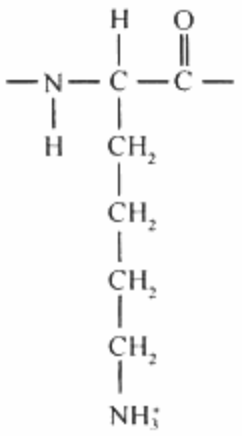
氨基酸家族

常见的氨基酸根据其侧链分为：酸性碱性不带电极性非极性 20 种氨基酸给出了 3 个字母或 1 个字母的缩写。如：丙氨酸＝Ala＝A

基本侧链

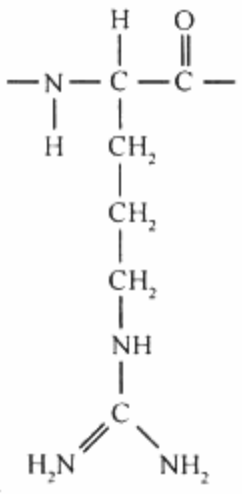
赖氨酸

(Lys, 或 K)



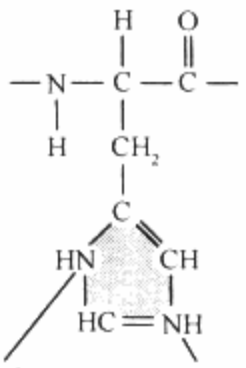
精氨酸

(Arg, 或 R)



组氨酸

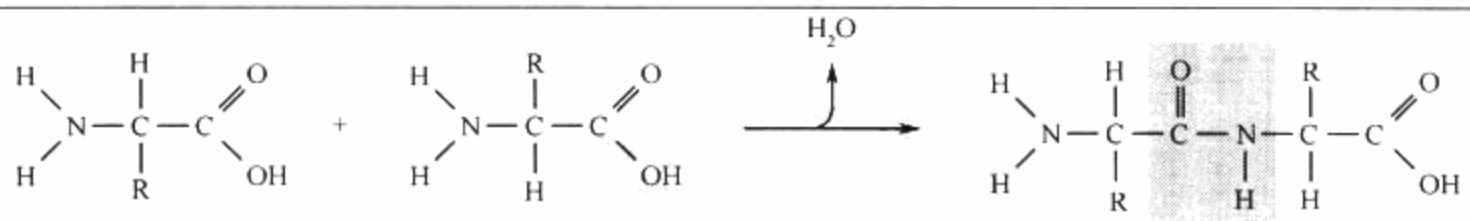
(His, 或 H)



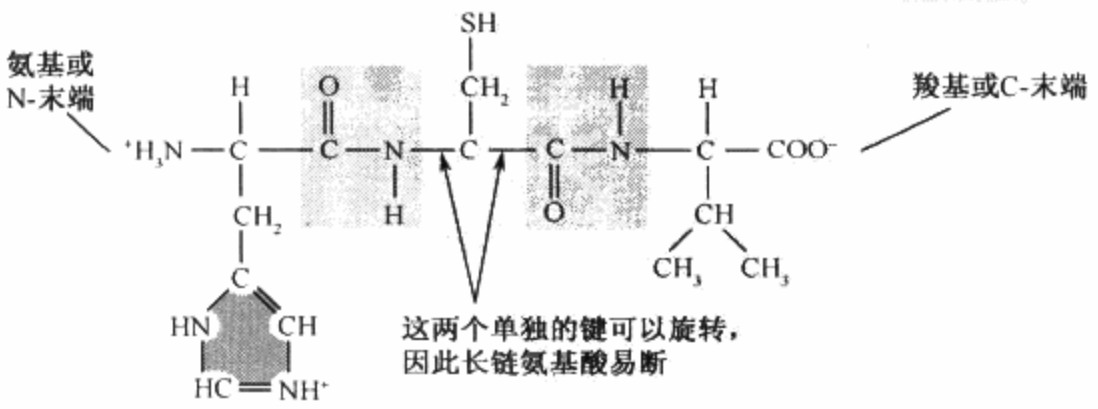
肽键

氨基酸通常通过被叫做肽键的氨基联接相连

肽键：灰色标出的四个原子位于同一个硬性平面上。围绕 C—N 键没有旋转。

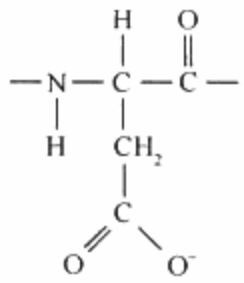


多聚氨基酸的长链蛋白通过钛键连接氨基酸。通常 N-端写在左边，这个三聚肽的序列为组氨酸-半胱氨酸-缬氨酸

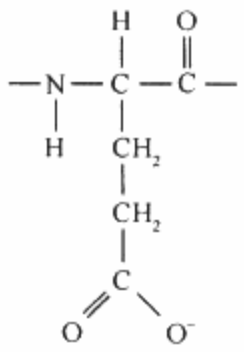


氨基侧链

天冬氨酸
(Asp 或 D)

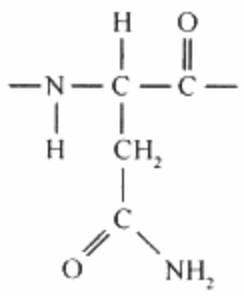


谷氨酸
(Glu 或 E)

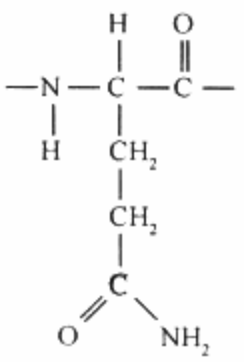


不带电的极性侧链

天冬酰胺
(Asn 或 N)

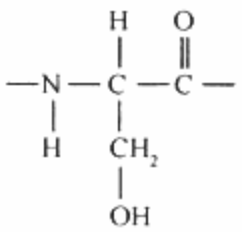


谷氨酰胺
(Gln 或 Q)

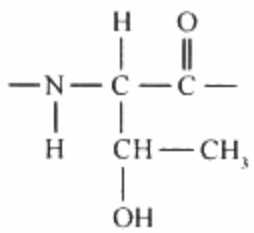


虽然在pH中性时酰胺N不带电，但它是极性的

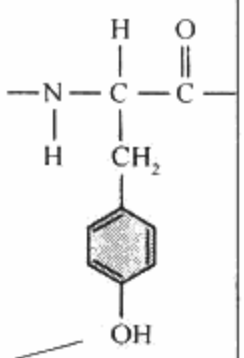
丝氨酸
(Ser 或 S)



苏氨酸
(Thr 或 T)



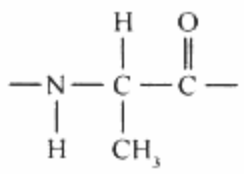
酪氨酸
(Tyr 或 Y)



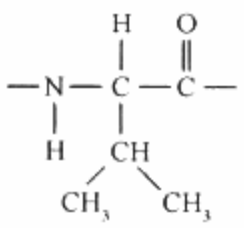
—OH基是极性的

非极性侧链

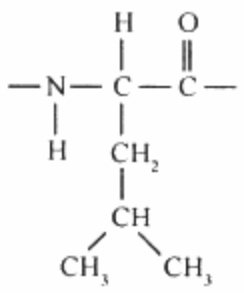
丙氨酸
(Ala 或 A)



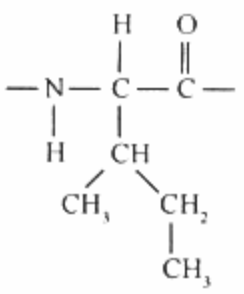
缬氨酸
(Val 或 V)



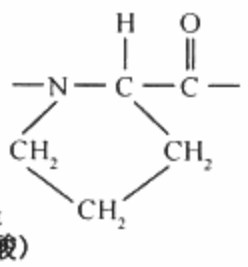
亮氨酸
(Leu 或 L)



异亮氨酸
(Ile 或 I)

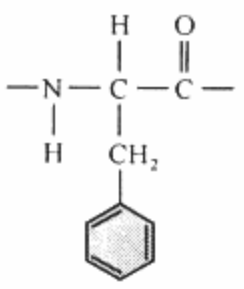


脯氨酸
(Pro 或 P)

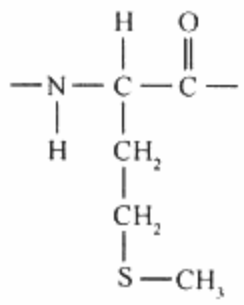


(实际是亚氨基酸)

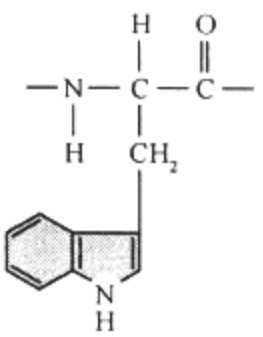
苯丙氨酸
(Phe 或 F)



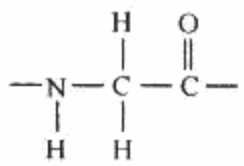
甲硫氨酸
(Met 或 M)



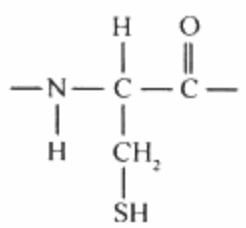
色氨酸
(Trp 或 W)



甘氨酸
(Gly 或 G)



半胱氨酸
(Cys 或 C)



蛋白质中两个半胱氨酸侧链可形成双硫键.

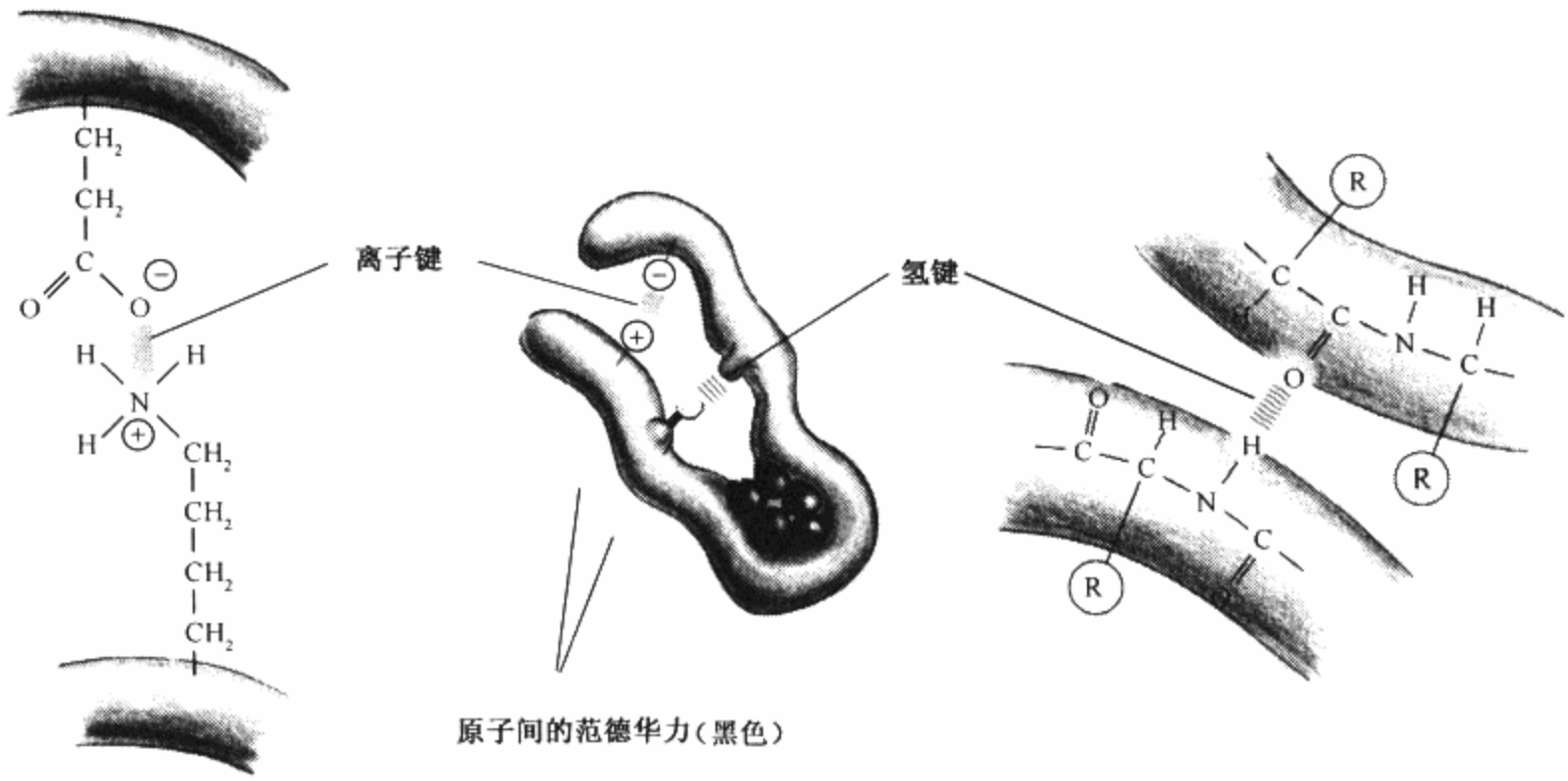


图 3-5　三种非共价键帮助蛋白质折叠

虽然一个这样的键是非常弱的，但是许多这种键一起可以形成一个强的键的排列，如图所示。R 代表一个氨基酸的侧链

第 4 种弱的作用力对蛋白质的结构形成也有非常重要的影响。如第 2 章所述，包括氨基酸的非极性侧链在内的疏水性分子，在水环境中为了减小其对水分子形成的氢键网络的破坏性作用，有聚集在一起的趋势（专题 2-2）。因此，决定蛋白质折叠的一个很重要的影响因素是极性和非极性氨基酸的分布。蛋白质中的非极性（疏水的）侧链，属于这一类的氨基酸，如苯丙氨酸、亮氨酸、缬氨酸和色氨酸，在分子内部易于成簇（正如疏水的油滴在水中聚集形成一个大的液滴）。这可以使它们免受细胞内水分子的影响。相反地，极性侧链，属于这一类的氨基酸，如精氨酸、谷氨酸和组氨酸，易于分布在分子外部，因为它们可以与水分子和其他极性分子形成氢键（图 3-6）。当极性氨基酸埋在蛋白质内部时，它们通常与其他极性氨基酸或者多肽链骨架形成氢键（图 3-7）。

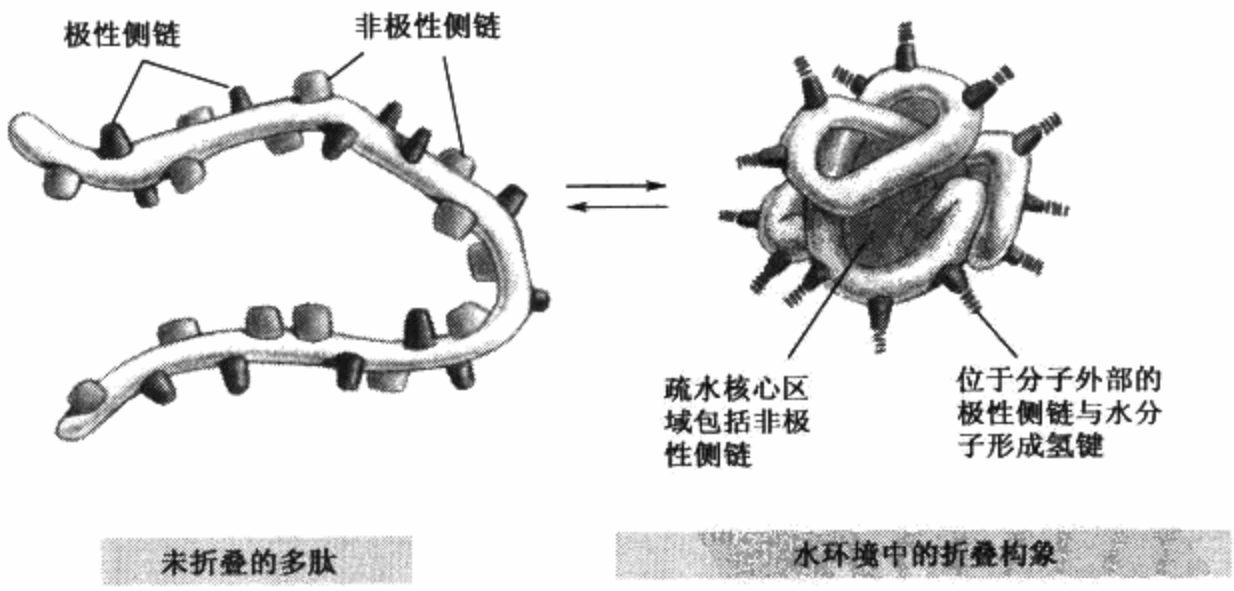


图 3-6　蛋白质如何折叠成一个紧密的构象

极性氨基酸侧链堆积在蛋白质外部，它们可以与水分子相互作用；非极性氨基酸侧链埋在蛋白质内部，形成一个紧密的疏水核心，可以远离水分子。在这个示意图中的蛋白质只包含 30 个氨基酸

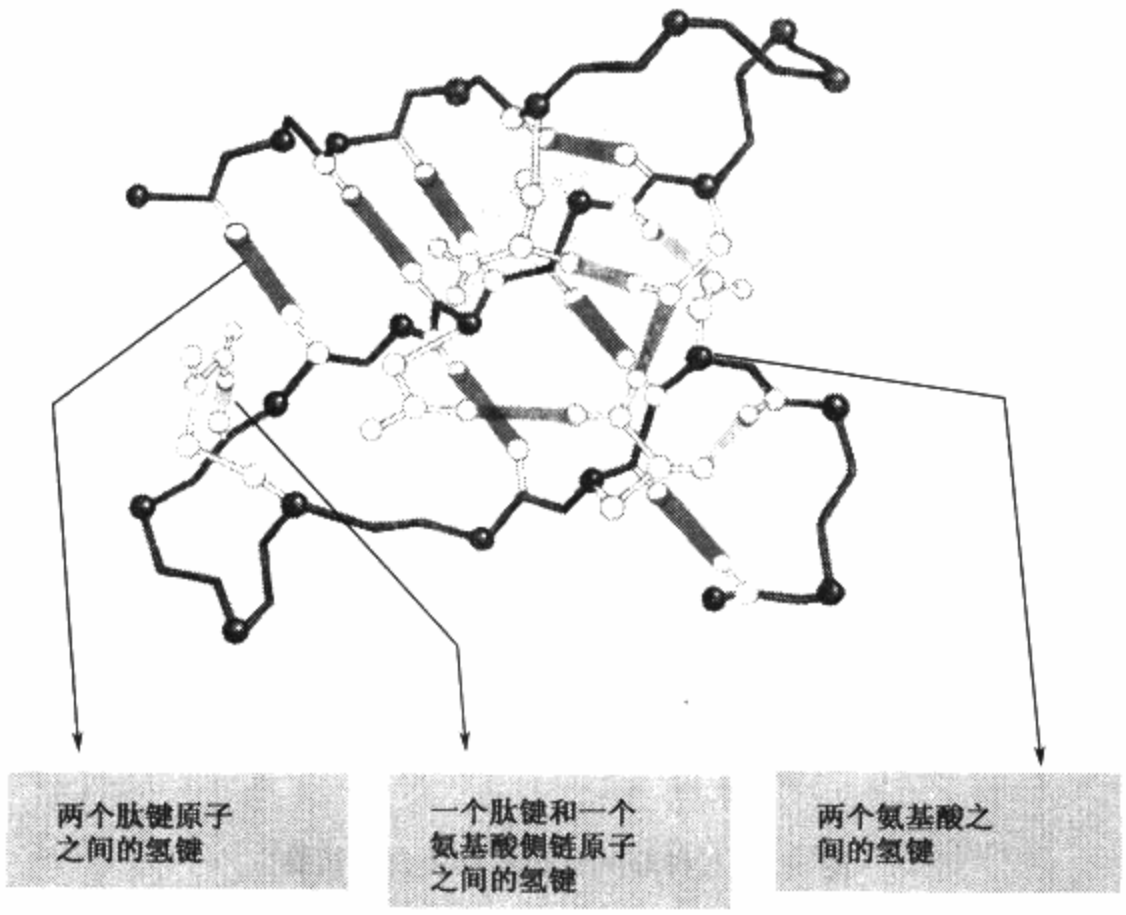


图 3-7　蛋白质分子中的氢键

折叠多肽链的相邻区域可以形成大量的氢键来帮助稳定其三维结构。图中所示的蛋白质是溶菌酶的一部分，可能形成氢键的三种情况被不同的颜色标记

蛋白质折叠为能量最低的构象

因为这些作用力的共同作用，每一种蛋白质有一个由其氨基酸序列决定的特定的三维结构。一条多肽链最终折叠形成的结构或**构象（conformation）**，通常其自由能最小。运用高度纯化的蛋白质可以在试管中研究蛋白质的折叠。特殊的溶剂处理蛋白质，可以使其去折叠，或者变性（denatured），它破坏了使多肽链折叠的非共价相互作用。这个处理使蛋白质成为一个松散的多肽链而失去了其天然的结构。当去掉变性剂，蛋白质可以自发地再折叠，或者复性（renature）为原来的构象（图 3-8）。这表明决定蛋白质特定三维结构的全部信息存在于其氨基酸序列中。

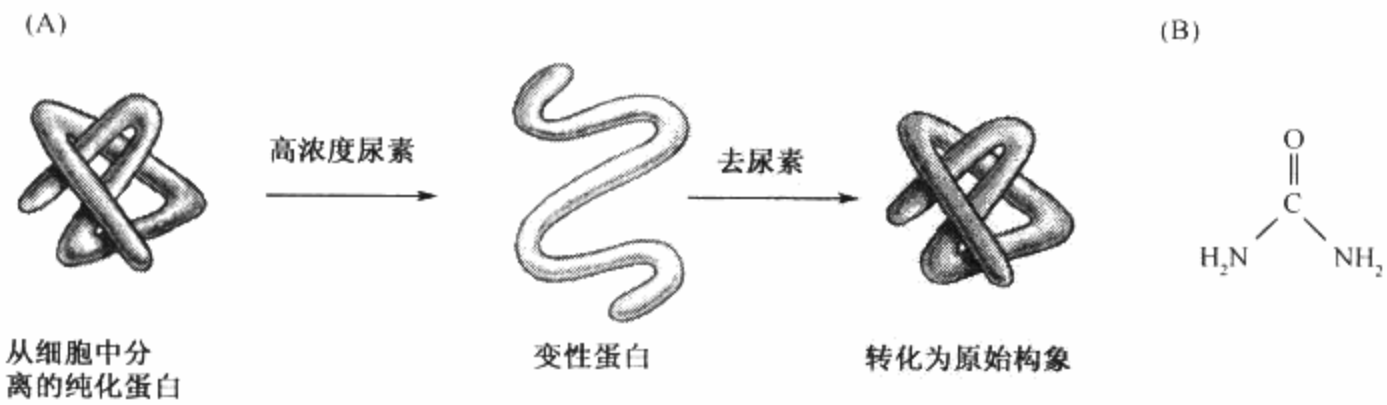


图 3-8　变性蛋白质的重折叠

（A）这个实验揭示了蛋白质的结构是由其氨基酸序列决定。（B）尿素的结构。尿素在水中溶解度很高，在高浓度下可以使蛋白质变性，每 6 个水分子结合一个尿素分子

每个蛋白质通常折叠为一个稳定的构象。但是，在细胞中，当蛋白质与其他分子相互作用时，其构象通常会发生轻微的改变。随后可以看到，这些结构上的变化对于蛋白质的功能是非常重要的。

虽然蛋白质可以无须帮助而折叠成为正确的构象，但是在生活细胞中，蛋白质折叠通常需要一类特殊称做分子伴侣（molecular chaperon）的蛋白质的帮助。这些蛋白质可以α蛋白发生相互作用而导致蛋白质聚集沉淀。但是，蛋白质最终的三维结构仍然是由其氨基酸序列决定的，分子伴侣只不过是帮助了折叠过程更加稳定。

蛋白质有多种多样的结构，它们的长度通常在50～200个氨基酸之间。大的蛋白质通常包含几个独立的蛋白质结构域（protein domain）——结构单位，其折叠或多或少是相互独立的。蛋白质的详细的结构描述是非常复杂的，为简便起见，蛋白质的结构可以采用不同的方式来描述，每一种强调的结构特点不同。

专题3-2显示了4种不同的描述SH2结构域（真核细胞中有重要功能）的方式。它由100个氨基酸组成，结构显示（A）多肽骨架模型、（B）丝带模型 、（C）线状模型。它包含了氨基酸侧链信息和（D）空间填充模型。三个水平箭头中的任一个指示蛋白质位于不同的方向。图像的颜色显示了多肽链从N端 到C端。

专题3-2显示了蛋白质的构象是非常复杂的，对于一个和SH2结构域一样小的结构也是如此。描述蛋白质结构也可以简单化，因为蛋白质都是由一些共同的结构基序组成，我们接下来将会讨论。

α螺旋和β折叠是常见的折叠模式

当比较许多不同蛋白质的三维结构时，可以清楚地看到，虽然每种蛋白质的总体构象是不同的，有两种规则的折叠模式通常存在。这两种模式是50年前在研究头发和丝的时候发现的。第一种折叠模式叫做**α螺旋**，它是在α角蛋白（α-keratin）中发现的，它在皮肤和其衍生物，如头发、指甲和角中大量存在。在发现α螺旋的同一年，在蚕丝蛋白（丝绸的主要成分）中发现了第二种折叠结构，**β折叠**。这两种折叠模式是很常见的，因为它们是由多肽链骨架上的N—H和C═O形成的氢键构成，并不涉及氨基酸的侧链。在每种情况下，多肽链形成一个规则的重复的构象。在蛋白质丝带模型中通常会指出这两种构象和其缩写，如图3-9所示。

很多蛋白质的核心包含大量的β折叠。在图3-10中，相邻的走向相同的多肽链（平行链）可以形成，或者一条多肽链来回折叠，使每个部分与其相邻部分走向相反（反平行链），也可以形成β折叠。两种类型的β折叠都会产生一个非常刚性的结构，它们由连接相邻链的氢键稳定结构［图3-9（D)］。

一条多肽链自我扭曲成一个刚性圆柱形时会产生α螺旋。每4个肽键形成一个由一个肽键的C═O和另一个肽键的N—H形成的氢键。一个规则的α螺旋每一圈包含3.6个氨基酸残基。注意在专题3-2中所示的蛋白质结构域包括2个α螺旋和几个β折叠。

定位于细胞膜上的蛋白质，如转运蛋白和受体蛋白通常富含大量短的α螺旋。在第10章中会讨论，跨膜蛋白都是以由非极性氨基酸组成的α螺旋来穿越细胞膜脂双层结构。亲水的多肽链骨架在α螺旋中都是以氢键自我相连，而且利用突出的非极性侧链来远离细胞膜的脂质疏水环境（图3-77)。

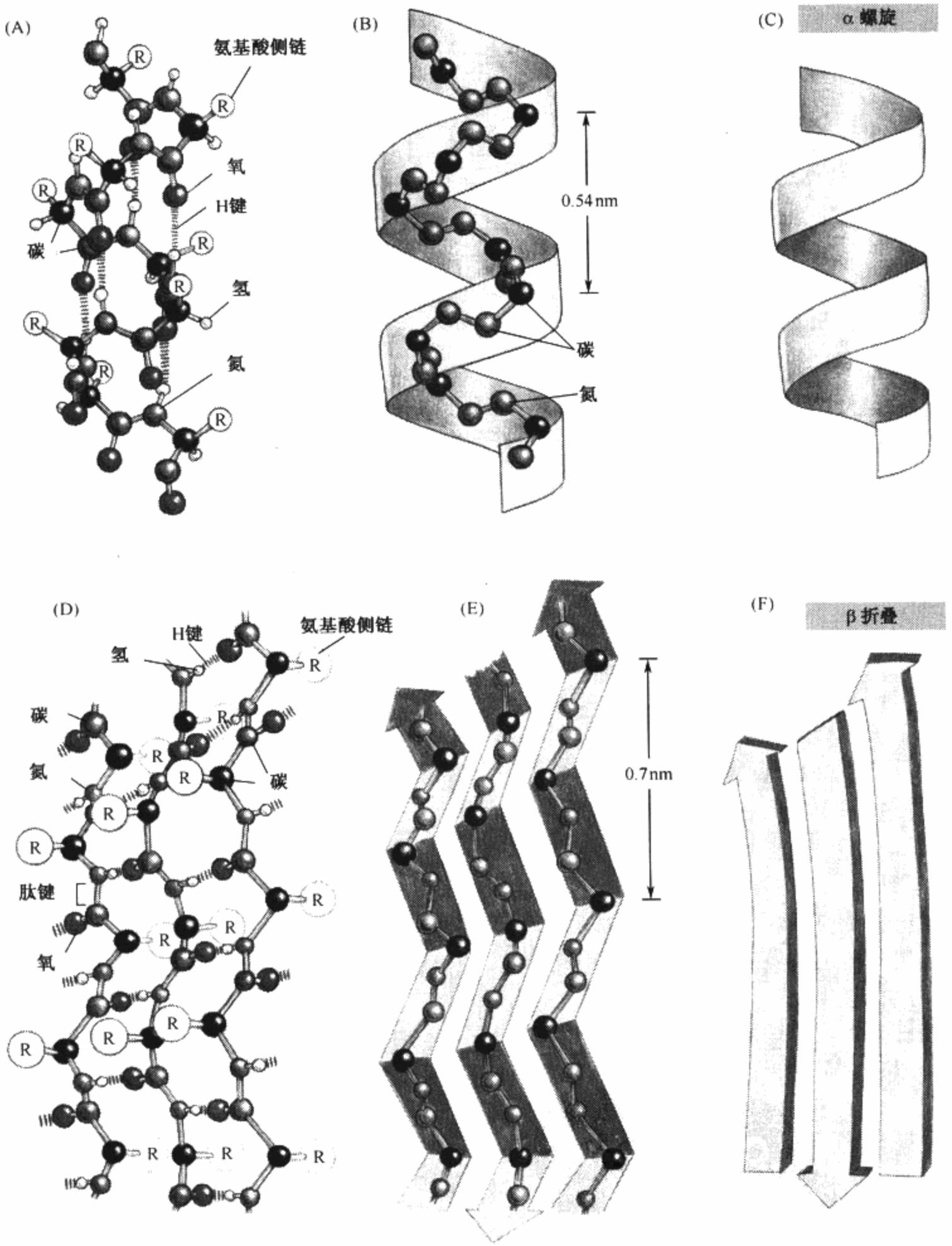
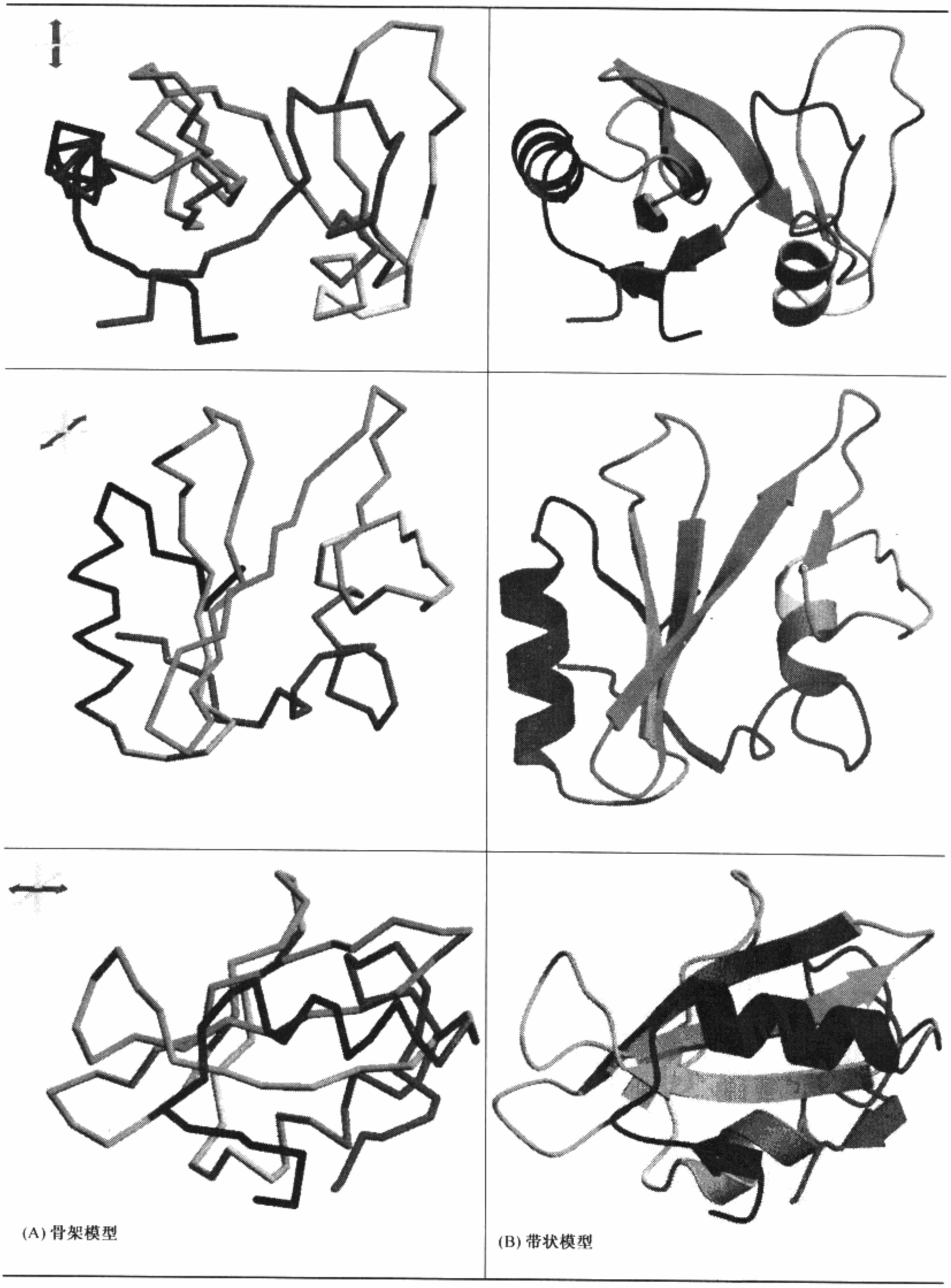


图 3-9　多肽链骨架上观察到的α螺旋和β折叠的规则构象

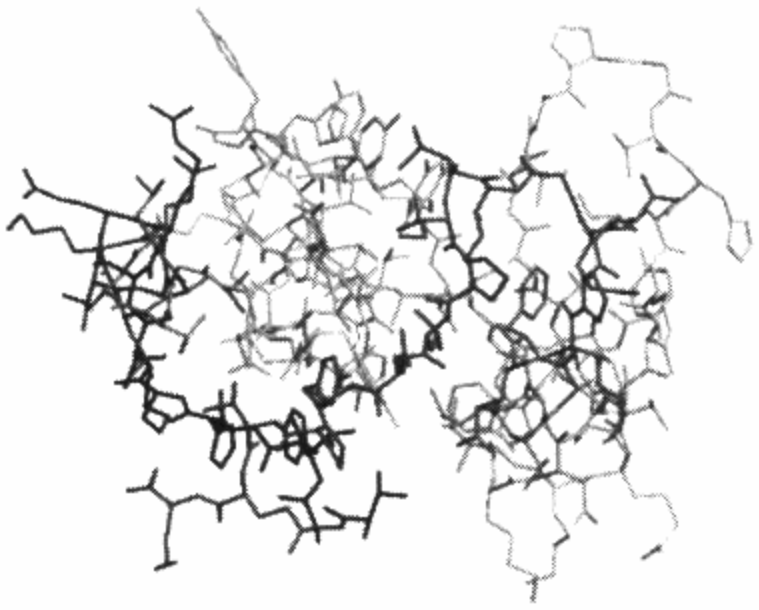
(A)、(B) 和 (C) α螺旋。每个肽键上的 N—H 和同一条肽链上它后面第 4 个残基上的 C═O 之间形成氢键。(D)、(E) 和 (F) β折叠。在这个例子中，相邻的肽链以相反的方式排列（反平行）。β折叠中的每个肽链（片）由两条链之间形成的氢键连接，每条链上的氨基酸侧链位于折叠平面的上面或下面。(A) 和 (D) 是多肽链骨架的所有原子，但是侧链被省去，以 R 表示。(B) 和 (E) 只显示了骨架原子，(C) 和 (F) 以一种简单的标记丝带模型来表示 α螺旋和β折叠

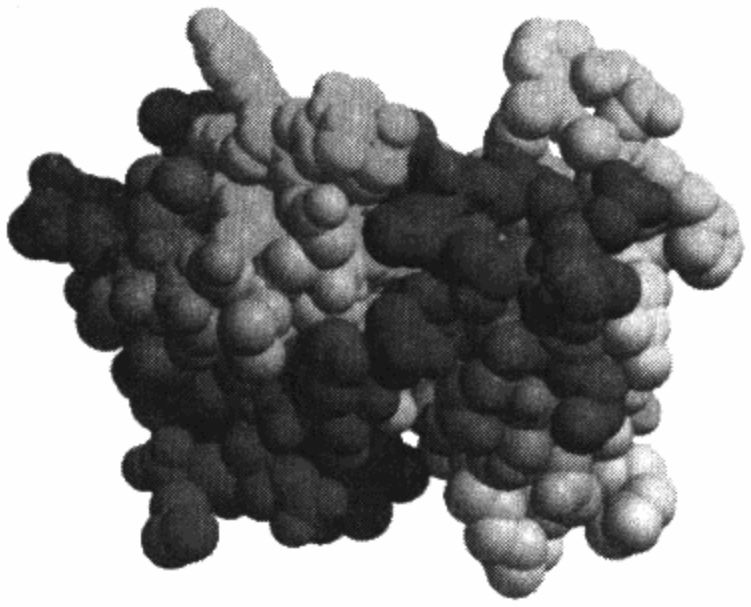
专题 3-2　显示了 4 种不同的描述 SH2 结构域（真核细胞中有重要功能）的方式

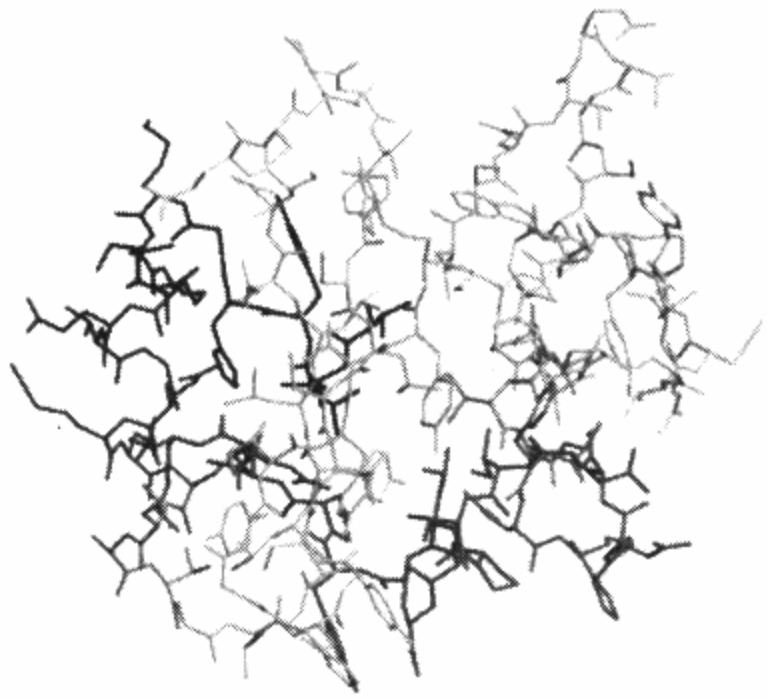


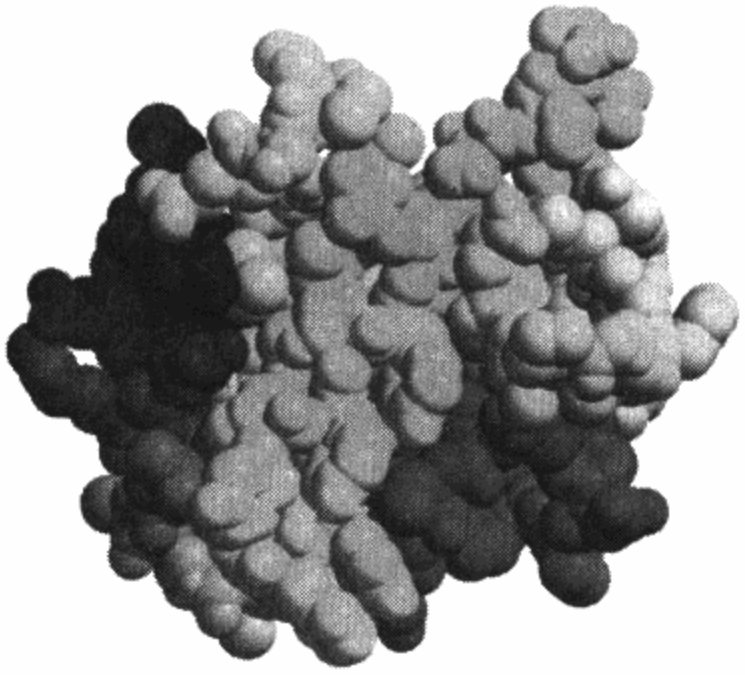
(A) 骨架模型

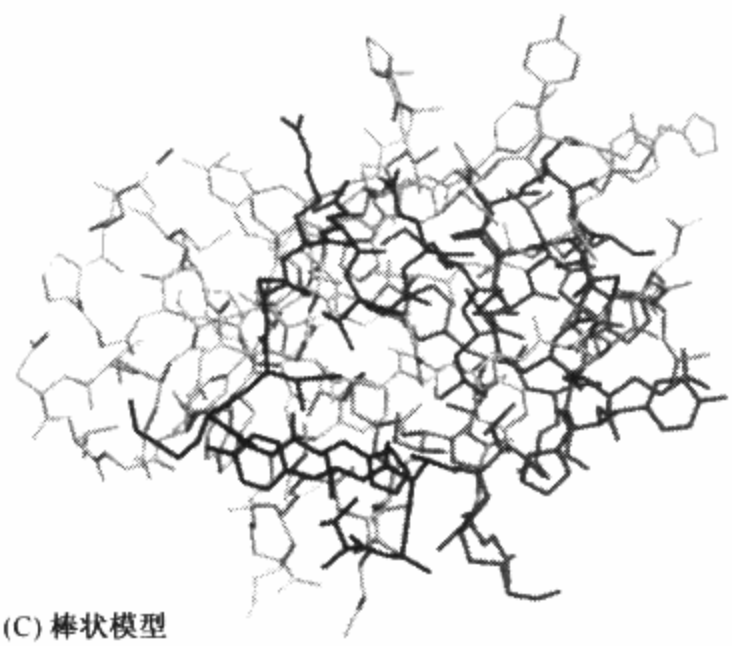
(B) 带状模型



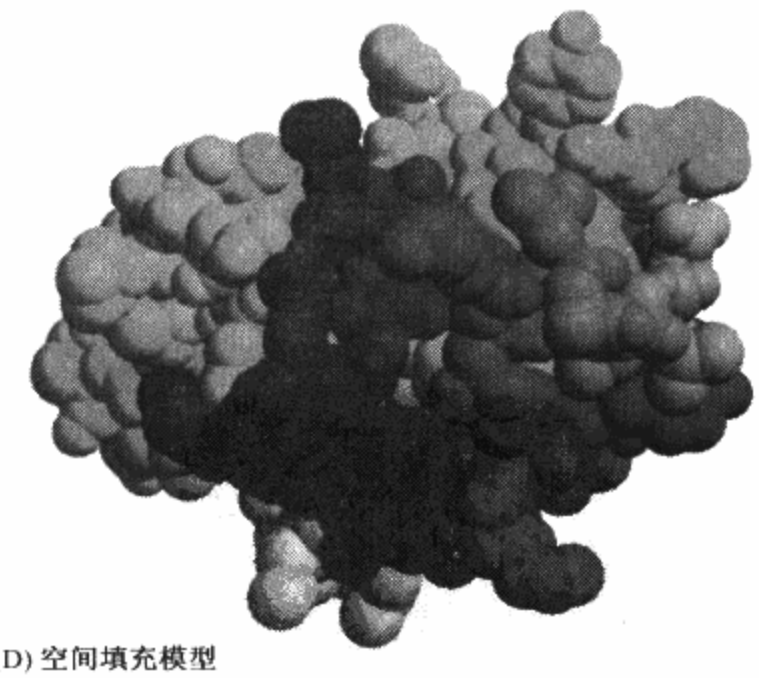








(C) 棒状模型



(D) 空间填充模型

在一些蛋白质中，α 螺旋相互缠绕形成一个特别稳定的结构 **coiled-coil**。当两个 α 螺旋（有时候 3 个）的大部分非极性氨基酸（疏水的）在同一边时易使这些侧链相互向内扭曲形成这种结构（图 3-11）。长的棒状 coiled-coil 结构为那些长的蛋白质提供了结构框架。α 角蛋白就是一个例子，它形成细胞内纤维来加固皮肤的外层及其附属成分，而肌球蛋白主要负责肌

肉的收缩。

结构域是蛋白结构的一个基本单位

即使一个小蛋白质，也是由成千上万个原子依靠一定方向的共价和非共价键精确连接起来的。如果缺少三维图像我们很难描述这样一个复杂的结构。因此，我们使用了许多图像的和基于计算机的辅助工具。

生物学家把蛋白质的结构组织分为 4 个层次。蛋白质的**一级结构**是指其氨基酸序列。蛋白质的**二级结构**是指伸展的多肽链形成的 α 螺旋和 β 折叠。蛋白质的**三级结构**指一条多肽链的三维结构。如果一个蛋白质分子是由多条多肽链形成的复合物，那么这个蛋白质完整的结构称为蛋白质的**四级结构**。

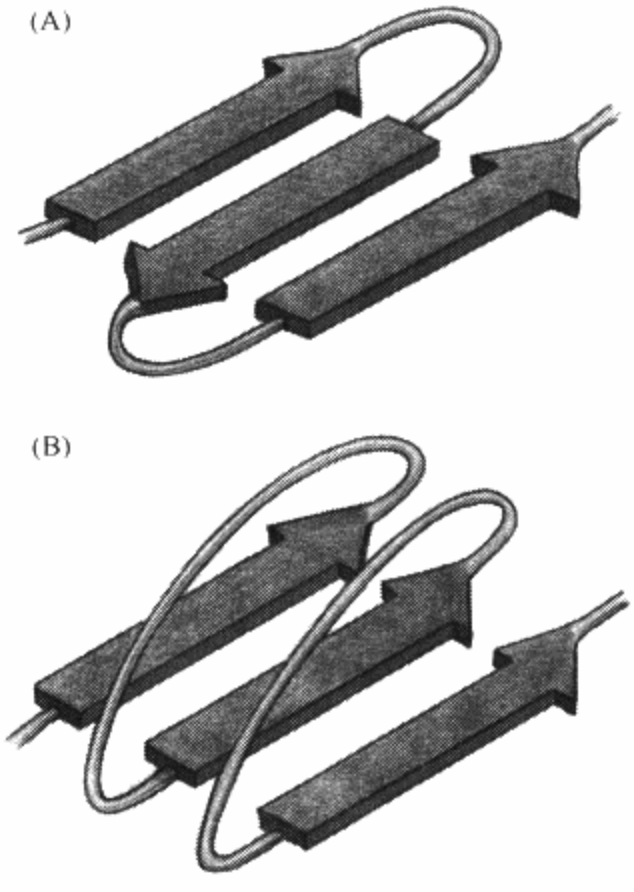


图 3-10　两种类型的 β 折叠结构

（A）反平行 β 折叠［图 3-9（D)］（B）平行 β 折叠。这两种结构在蛋白质中很常见

对蛋白质的构象、功能和进化的研究表明这 4 个层次之外还存在一种独立的结构单位有着重要的作用——**蛋白质结构域**，它是多肽链上的一部分独立的折叠而成一个紧密的稳定的亚结构。一个结构域通常包含 40～350 个氨基酸，它是许多大的蛋白质的模式单位。一个蛋白质中不同的结构域通常具有不同的功能。图 3-12 给出一个例子——SRC 蛋白激酶，它在脊椎动物细胞中起信号传递的作用。它包含 4 个结构域：SH2 和 SH3 结构域起调节功能，另外两个结构域负责激酶的催化活性。在这一章的后面，我们会回到这个蛋白质来解释蛋白质怎样形成分子开关在细胞中传递信息。

最小的蛋白质分子只有一个结构域，而大的蛋白质可能包含几十个结构域，它们之间通过短的、相对无结构的多肽链相连。图 3-13 显示了 3 个不同的蛋白质结构域的丝带模型。这些例子表明，一个结构域的核心可以由 α 螺旋、β 折叠或者这两种基本折叠单元的不同组合构成。每一种组合也称作一个蛋白质折叠方式。目前为止，已知构象的 1 万个蛋白质中大约鉴定出有 1000 种不同的蛋白质折叠方式。

可能的多肽链中只有少数是有用的

由于 20 种氨基酸的化学性质各不相同，理论上，每个氨基酸在一条多肽链上的每个位置都有可能出现，那么一个由 4 个氨基酸组成的蛋白质可能有 $20\times20\times20\times20=160\ 000$ 条可能的多肽链。对于一个 N 个氨基酸长的蛋白质，可能有 $2N$ 条不同的多肽链。一个典型的含有 300 个氨基酸的蛋白质，理论上可以产生大于 10^{390}（20^{300}）条不同的多肽链。这是一个巨大的数目，产生每一个分子需要比宇宙现存的原子多得多的原子数。

这些可能的多肽链中只有极小一部分可以形成一个单一的稳定的三维构象，据估计，少于 1/10 亿。这样数量巨大的可能的蛋白质分子可以形成很多稳定性大致相同的构象，每一种构象具有不同的化学性质。但是实际上，细胞中的所有蛋白质的稳定构象是唯一的。这是为什么呢？答案在于自然选择。一个蛋白质如果拥有多变的结构和生物化学特性并不能帮助含有此蛋白质的细胞的生存。详细内容参见图 3-14、图 3-15 和图 3-16。

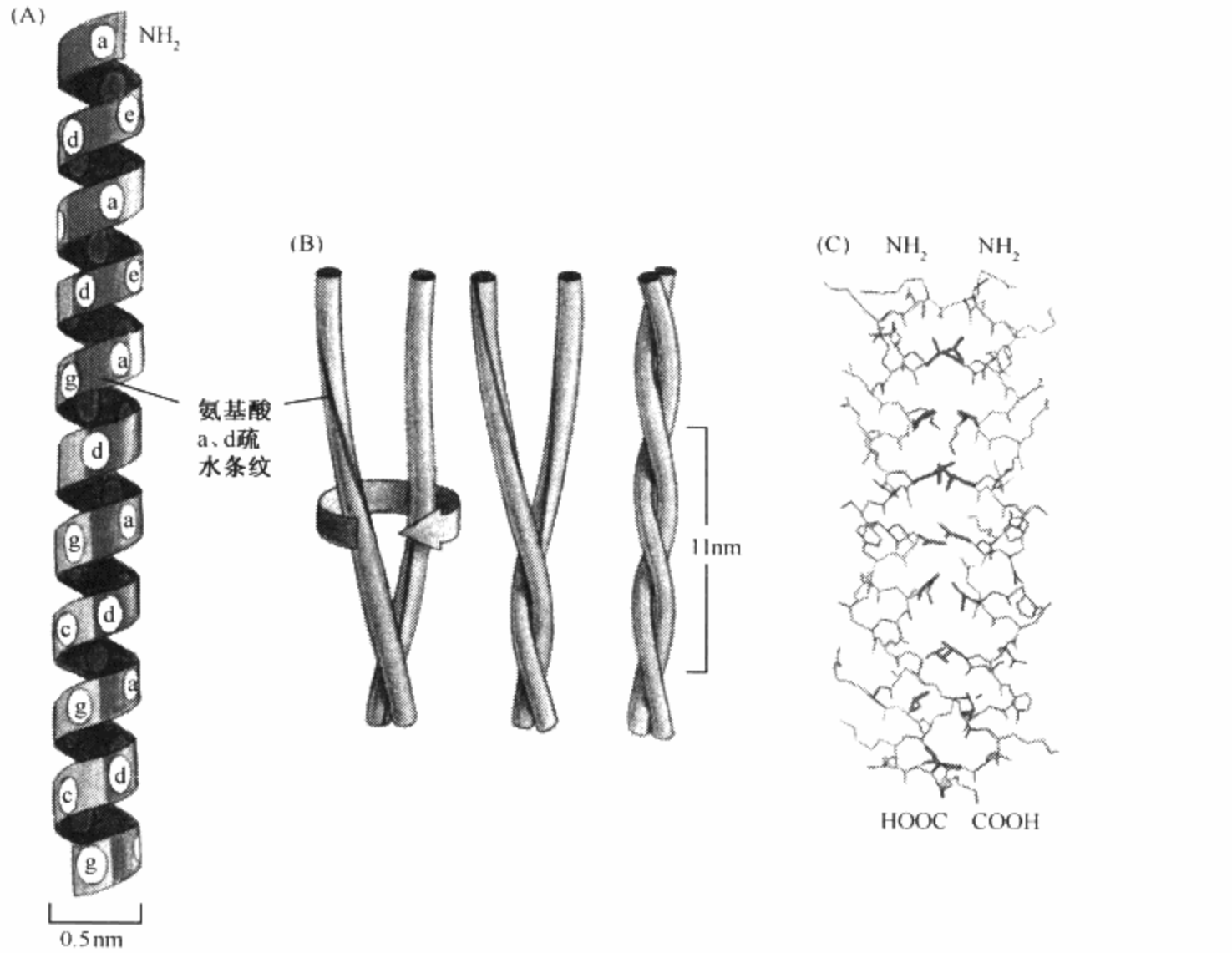


图 3-11　coiled-coil 的结构

（A）一个单独的 α 螺旋，连续的氨基酸以 7 个字母（abcdefg）标记，（从下到上）。这种排列方式的“a”和“d”氨基酸位于圆柱表面距离非常近，形成一个“长条”（红色）围绕在 α 螺旋周围。形成 coiled-coil 结构的蛋白质都有非极性氨基酸位于“a”和“d”。如（B）所示，两个 α 螺旋相互缠绕，以一个 α 螺旋的非极性侧链和另一个 α 螺旋的非极性侧链相互作用，剩余的亲水性氨基酸的侧链暴露在水环境中。（C）通过 X 射线晶体衍射技术得到的 coiled-coil 结构的原子结构。红色的侧链是非极性的

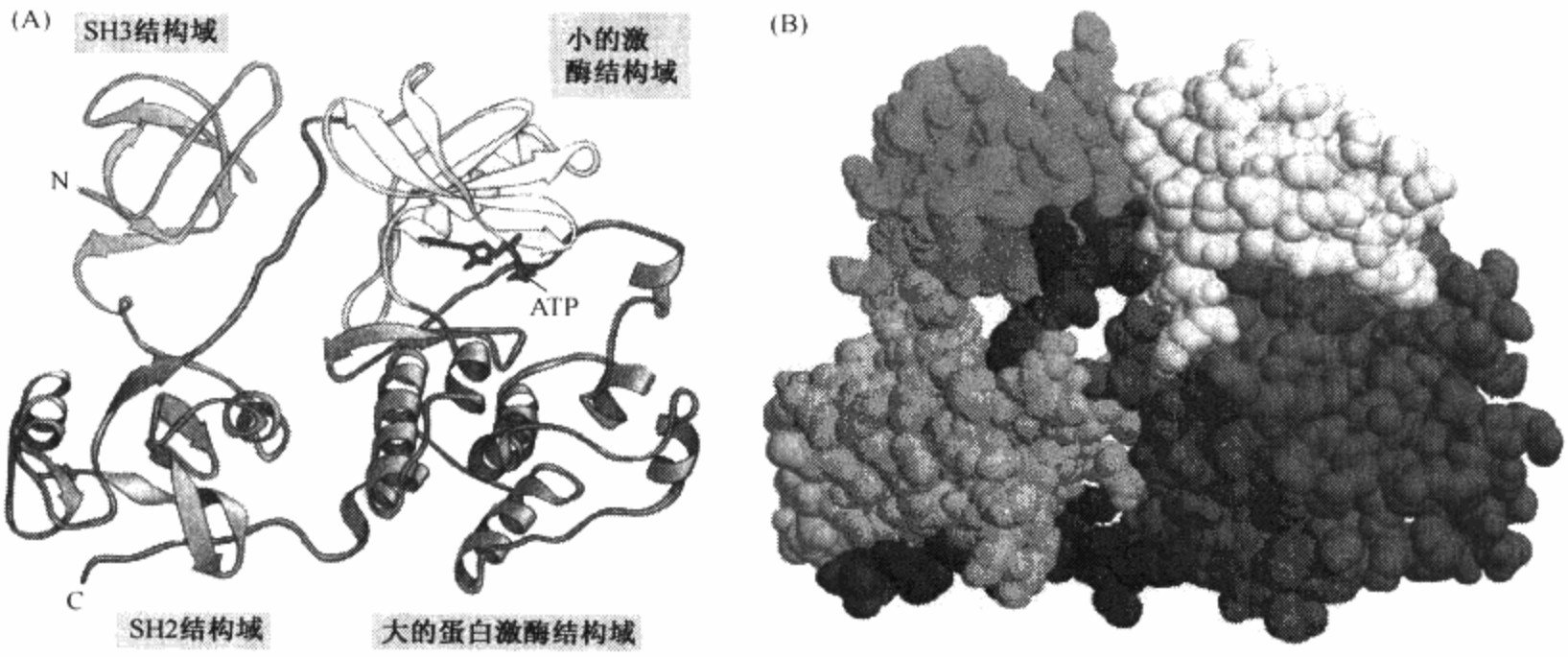


图 3-12　一个蛋白质由 4 个结构域组成

在图中的 Src 蛋白中，两个结构域形成蛋白质的激酶部分，而 SH2 和 SH3 结构域执行调节功能。（A）丝带模型，红色的是 ATP 底物。（B）空间占有模型，红色表示底物 ATP。注意到结合 ATP 的位点位于构成激酶的两个结构域交界处。SH2 结构域的详细结构在专题 3-2 中

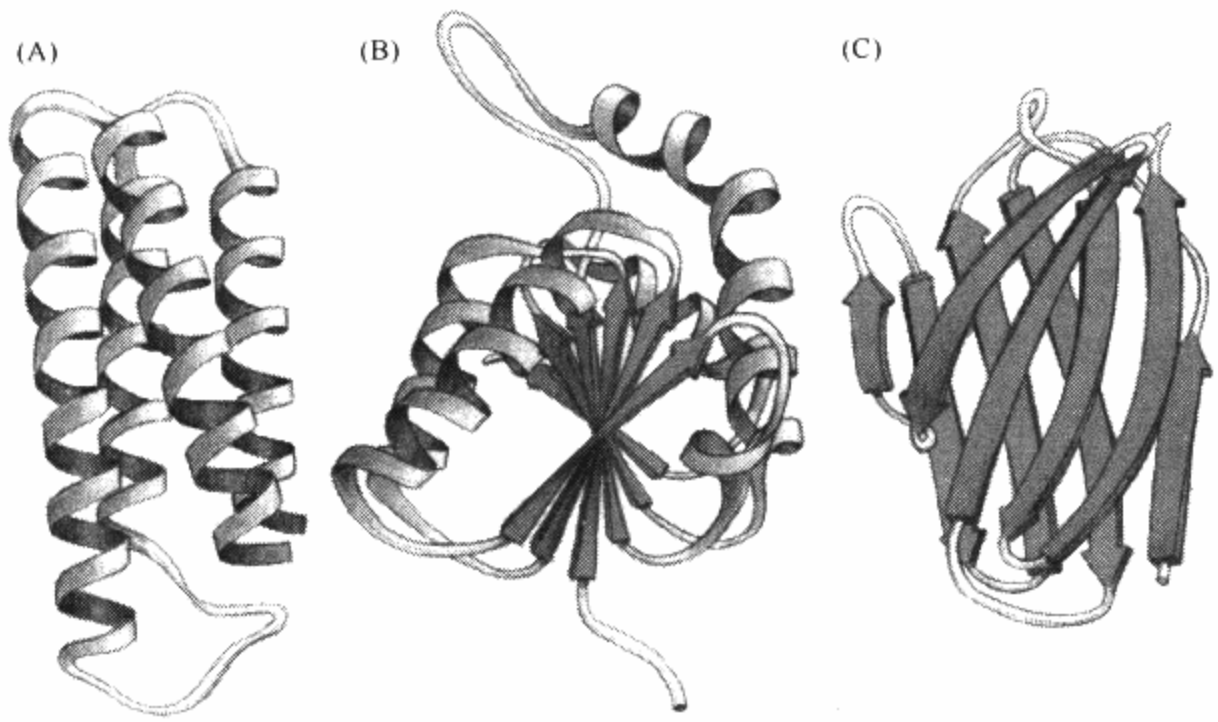


图 3-13　三种不同蛋白质的丝带模型

（A）细胞色素 b_{562}，一种单结构域蛋白质，参与线粒体内的电子传递。该蛋白质几乎全由 α 螺旋组成。（B）乳糖脱氢酶的 NAD 结合结构域，由 α 螺旋和 β 折叠构成。（C）免疫球蛋白（抗体）轻链的可变的结构域，由 β 折叠形成的三明治构成。这些例子中，绿色代表 α 螺旋，红色箭头代表 β 折叠

注意在整个结构域中多肽链经常会前后折叠。只在蛋白质表面有快速的转弯。突出的环结构（黄色）通常是其他分子的结合位点

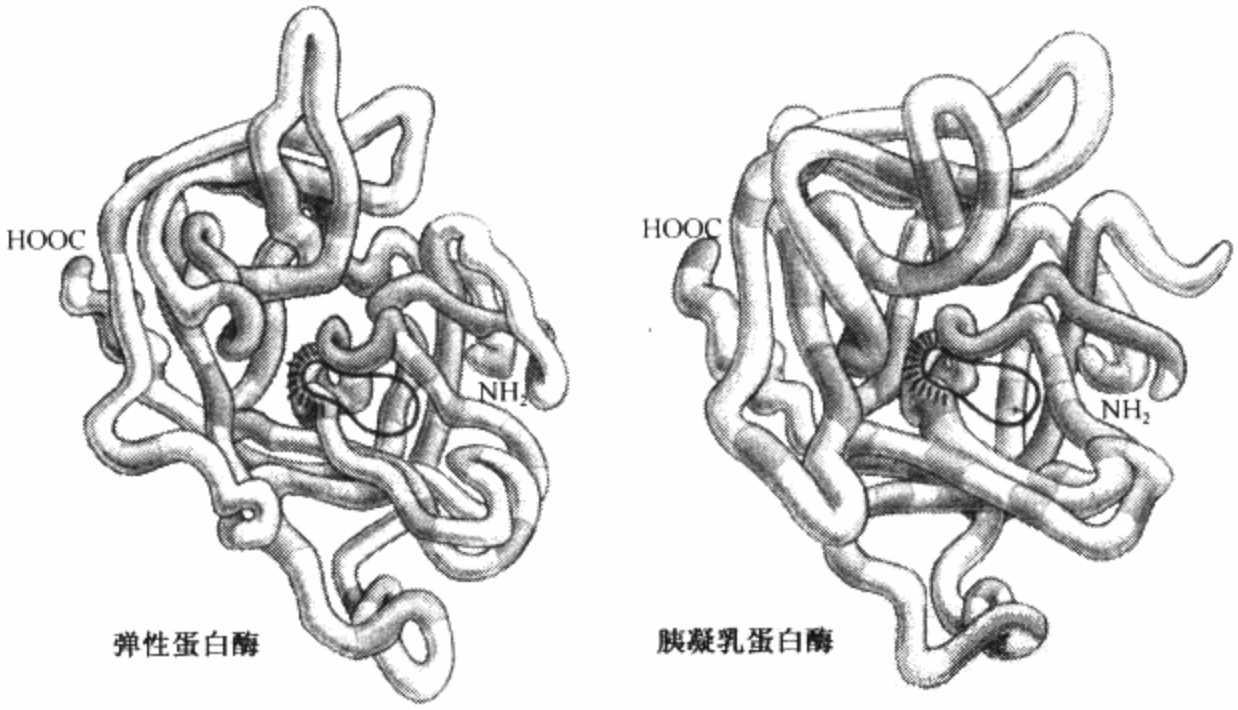


图 3-14　比较两种丝氨酸蛋白酶的构象

弹性蛋白酶和胰凝乳蛋白酶骨架的构象。绿色代表的是两条多肽链相似的氨基酸序列，虽然序列相似性很少，但是它们的构象几乎完全相同。每个酶的活性中心由红色的环表示；这个位置是蛋白质肽键和底物结合的地方并被水解。丝氨酸蛋白酶的名字来源于丝氨酸，它的侧链是酶的活性中心，参与到切割反应

蛋白质可以形成有限的折叠模式

我们对细胞知识的了解的增长速度是惊人的。1950 年，我们还不知道一条多肽链的氨基酸序列，很多人甚至怀疑蛋白质中的氨基酸是精确排列的。1960 年的时候，已经用 X 射线衍

射解出了第一个蛋白质的三维结构。现在我们可以根据编码基因的 DNA 序列来了解成千上万的蛋白质序列。我们期望下一个技术又是什么呢?

从一个基因的序列到表达其所编码的蛋白质不再是一个难题。利用 DNA 克隆和遗传工程技术（见第 8 章），这个工作是很常见的。但是了解一个蛋白质的结构却仍是一个难题。基于 X 射线衍射的标准技术需要蛋白质在一个合适的环境中形成一个大的、规则排列的晶状阵列——蛋白质晶体。在这方面每一个蛋白质的性质是很不相同的。只有通过不断的试验、摸索，花费很长的时间才可能成功得到晶体。

膜蛋白和包含很多组分的大的蛋白质复合物是最难于结晶的，这也是为什么这本书中只展示了很少一部分这类蛋白质的结构的原因。因此，大的蛋白质一般通过解决其每一个结构域的构象来进行分析：通过结晶单独的结构域，然后用 X 射线对晶体轰击，或者通过核磁共振技术（NMR）（见第 8 章）了解其各个结构域在溶液中的构象。通过结合 X 射线衍射和核磁共振技术，我们已经知道了上万种蛋白质的三维结构或构象。

通过仔细比较已知蛋白质的构象，结构生物学家（生物分子结构方面的专家）总结出蛋白质结构域的折叠模式是有限的，可能少于 2000 个。目前，还有 1000 种蛋白质折叠模式没有解决；因此我们已知的只是可能存在的蛋白质结构域折叠模式总数的一半。列出所有存在于活的有机体的蛋白质折叠模式的清单是我们研究的目标。

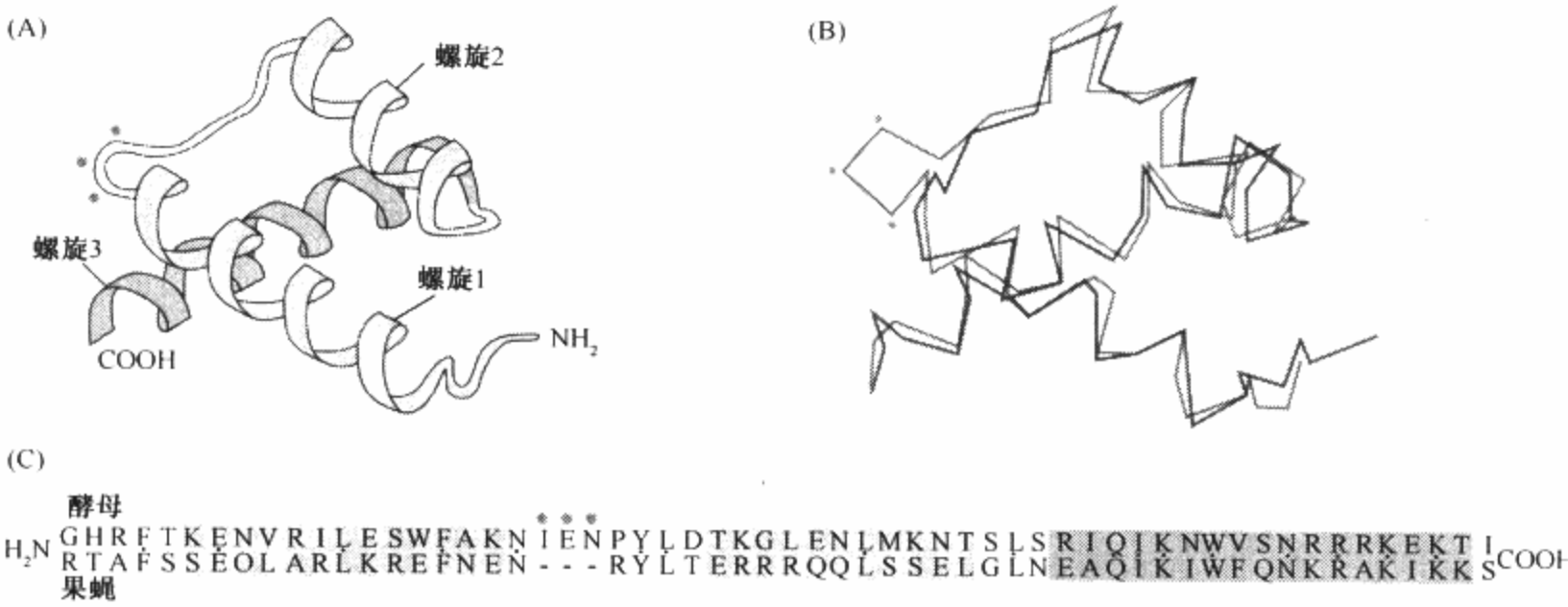


图 3-15　比较一类 DNA 结合结构域——同源结构域，比较的蛋白质来源于进化上相隔亿年的两个物种

（A）丝带模型所示的结构对于蛋白质是相同的。（B）α 碳原子的轨迹。所示的三维结构是由 X 射线衍射解出的酵母 α2 蛋白（绿色）和果蝇锯齿状蛋白（红色）。（C）比较在（A）图和（B）图中显示的蛋白质区域的氨基酸序列。黑点代表相同的氨基酸。橙色的点表示 α2 蛋白中有 3 个氨基酸插入（摘自 C. Wolberger et al.，*Cell* 67：517～528，1991）

同源序列搜索可以鉴定亲缘关系

现有的蛋白质序列数据库包含了约 500 000 个条目，随着越来越多的基因组被测序，大量编码蛋白质的基因被揭示，这个数字会飞速增长。强大的计算机搜索程序可以将新发现的蛋白质和整个数据库进行比对来寻找相似的蛋白质。同源蛋白质的定义是它们的基因从一个共同的古老的祖先进化而来，可以通过统计学上比较其氨基酸序列的相似性来鉴定它们。

由于数据库中存在大量的蛋白质数据，搜索程序可能发现很多非特异的配对，这样就造成了一个背景噪声，使得难以找到亲缘关系最近的序列。一般来说，两个蛋白质的序列相似

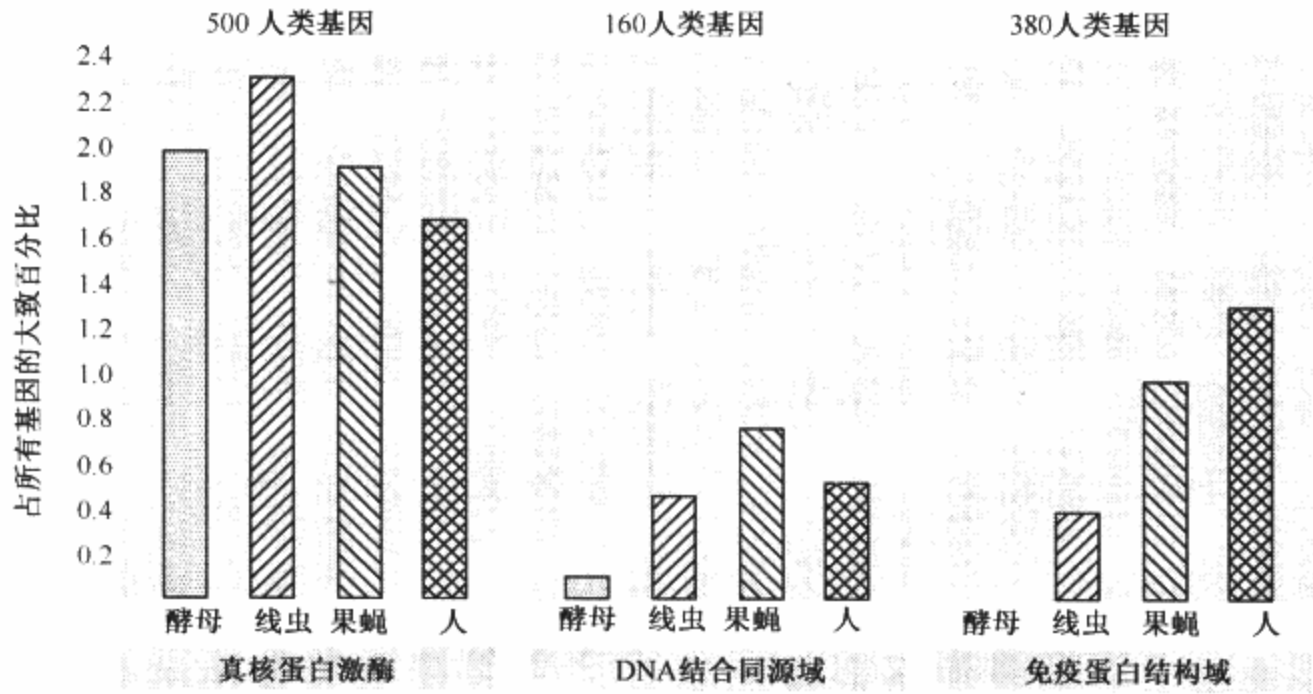


图 3-16　根据全基因组序列，所示的包含一个或多个拷贝特定结构域的所有基因的比例

注意所选的 3 个结构域中的免疫球蛋白结构域是最新被加入的，其相对含量在脊椎动物中还在增加。人类基因的数量是一个大概的估计

性至少要在 30%以上才能进行判断。但是，很多短的序列（“指纹”）代表了蛋白质已知的特殊的功能，可以用来寻找远源的同源（图 3-17）。

```
WYFGKITRRESERLL        GTFLVRESE                          - 信号序列
WYFGKITRRESERLLLNAENPRGTFLVRESETTKGAYCLSVSDFDNAKGL - 人类
W+F  + R+E+++LLL  ENPRGTFLVR SE     Y LSV D+++ +G  - 序列比对
WFFENVLRKEADKLLLAEENPEGTFLVRPSEHNPNGYSLSVKDWEDGRGY - 果蝇
1         10        20        30        40        50
```

图 3-17　通过短的标签序列来搜寻同源结构域

显示两个短的序列，分别为 15 个和 9 个氨基酸（绿色）可以用来在大的数据库中搜索蛋白质结构域，如 SH2 结构域。这里比较了人类和果蝇 Src 蛋白的 SH2 结构域 100 个氨基酸的前 50 个氨基酸（图 3-12）。通过计算机进行序列比对（黄箭头），人类和果蝇蛋白质之间的精确匹配由氨基酸的单字母简写来表示；相似的但不完全相同的氨基酸用＋表示，不配对的部分用空格表示。在此图中，一个或两个蛋白质都包含一个和绿色序列精确匹配的部分，比对的序列用红色表示

蛋白质的比对是非常重要的，因为结构相似也意味着功能上的相关。如果一个新的蛋白质序列与一个已知功能的蛋白质序列有同源性的话，可以省去很多实验的时间花费。如根据同源序列发现那些能使哺乳动物细胞变为癌细胞的基因是蛋白激酶。同样地，很多控制果蝇胚胎发育过程中形态建成的蛋白质很快的被发现是基因调控蛋白质。

计算机技术可以将氨基酸序列归类为已知的蛋白质折叠模式

我们知道有很多的方法可以产生很多三维结构相同的不同蛋白质，经过进化，随机的突变可以在不改变蛋白质构象的情况下改变氨基酸的序列。因此，结构生物学家现今的目标是解决自然界中蛋白质所有不同的折叠模式，并且开发计算机程序检测一个结构域中的氨基酸序列来鉴定这个结构域属于已知的哪种折叠方式。

一个称为“threading”的计算机技术可以用来将氨基酸序列归类为一个特定的蛋白质折

叠模式。对每一种已知的可能蛋白质折叠方式，计算机会搜索这段氨基酸序列最符合的结构。是否疏水氨基酸都处于蛋白质内部？是否有很强的形成 α-螺旋趋势的序列确实是 α-螺旋？等等。最匹配的得分最高，反映了预测的结构的稳定性。

在很多情况下，一段序列与一个特定的三维结构非常匹配，表明了这个结构域最可能的结构。另外一些情况下，已知的折叠方式都不太可能。运用 X 射线衍射和核磁共振技术，结构生物学家希望能快速地了解更多的蛋白质折叠方式，他们的目标是建立一个包含自然界存在的所有蛋白质结构的数据库。有了这个数据库，再加上一些计算机技术的改进，我们最终可能在知道一个氨基酸序列后快速地得到其大概的三维结构。

一些被称为模块的蛋白质结构域，形成了很多不同蛋白质的一部分

如前所述，大部分蛋白质是由一系列蛋白质结构域组成，结构域是蛋白质多肽链的不同区域独立折叠形成的紧密结构。这些多结构域的蛋白质被认为是编码每个结构域的这些基因的偶然组合而形成的新的基因。在结构域的毗连处通常会形成新的结合表面，很多蛋白质结合小分子的功能性位点位于此（例子见图 3-12）。很多大的蛋白质清楚的显示是由已存在的结构域的重新组合形成的，这个进化过程叫做结构域洗牌（shuffling）（图 3-18）。

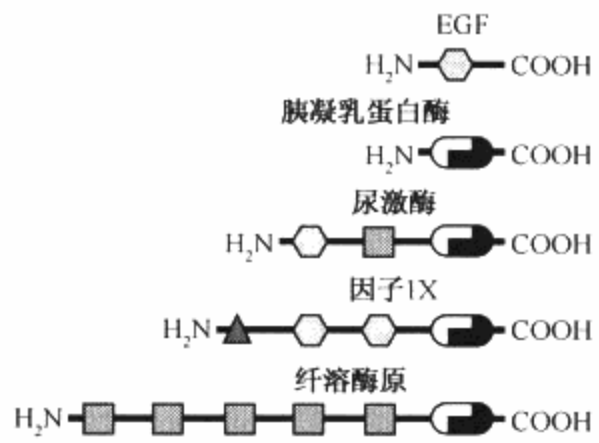


图 3-18　结构域洗牌

在蛋白质进化过程中会发生很多的蛋白质结构域洗牌。图中所示的形状和颜色相同的蛋白质结构域通常是同进化相关的。丝氨酸蛋白酶如胰凝乳蛋白酶由 2 个结构域组成（棕色）。所示的 3 个其余的蛋白酶，其调控和功能更加复杂，这两种蛋白酶结构域一个或多个与人类表皮生长因子（EGF，绿色）同源的结构域连接，或者钙离子结合结构域（黄色），或者含有 3 个内部二硫桥的“kringle”结构域（蓝色），图 3-14 有胰凝乳蛋白酶的解释

一些在进化中有特殊的运动性蛋白质结构叫做**蛋白质模块**，它们通常比结构域的平均大小要小（40～200 氨基酸），它们有特别的通用结构。这种蛋白质模块的一个典型代表就是 SH2 结构域，在专题 3-2 有说明。图 3-19 显示了其他一些蛋白质模块的结构。

每一个模块有一个由 β 折叠和一些多肽链突出部分（绿色）比较无规则的环构成的稳定的核心结构。环的结构是其他分子的结合位点，免疫球蛋白是一个非常典型的采取这样折叠方式的蛋白质，这也是形成抗体分子的基础（图 3-42）。基于 β 折叠的模块的进化是由于它为产生新的配体结合位点提供了一个方便的框架，其突出环的小的改变就可以改变其配体结合特性。

蛋白质模块的第二个特点是它们很容易和其他蛋白质整合，这也解释了它们的实用性。图 3-19 中的 6 个模块其中 5 个的 N 端和 C 端位于极性相反的位置。这种“线性”排列意味着编码这种模块分子的 DNA 分子经过随机的复制，这在基因组的进化中是很常见的（讨论见第 7 章），被加倍的模块分子可以连接形成伸展的结构，同种模块分子自身相连接或不同线性分子之间相互连接（图 3-20）。由很多模块分子构成的刚性伸展结构通常存在于细胞外基质分子和细胞表面受体蛋白质的胞外部分。其他包括 SH2 结构域和图 3-19 中的 kringle 在内的模块分子，形成一个“插销”形。在基因组重排后，这些模块分子通常插入另一个蛋白质的环区。

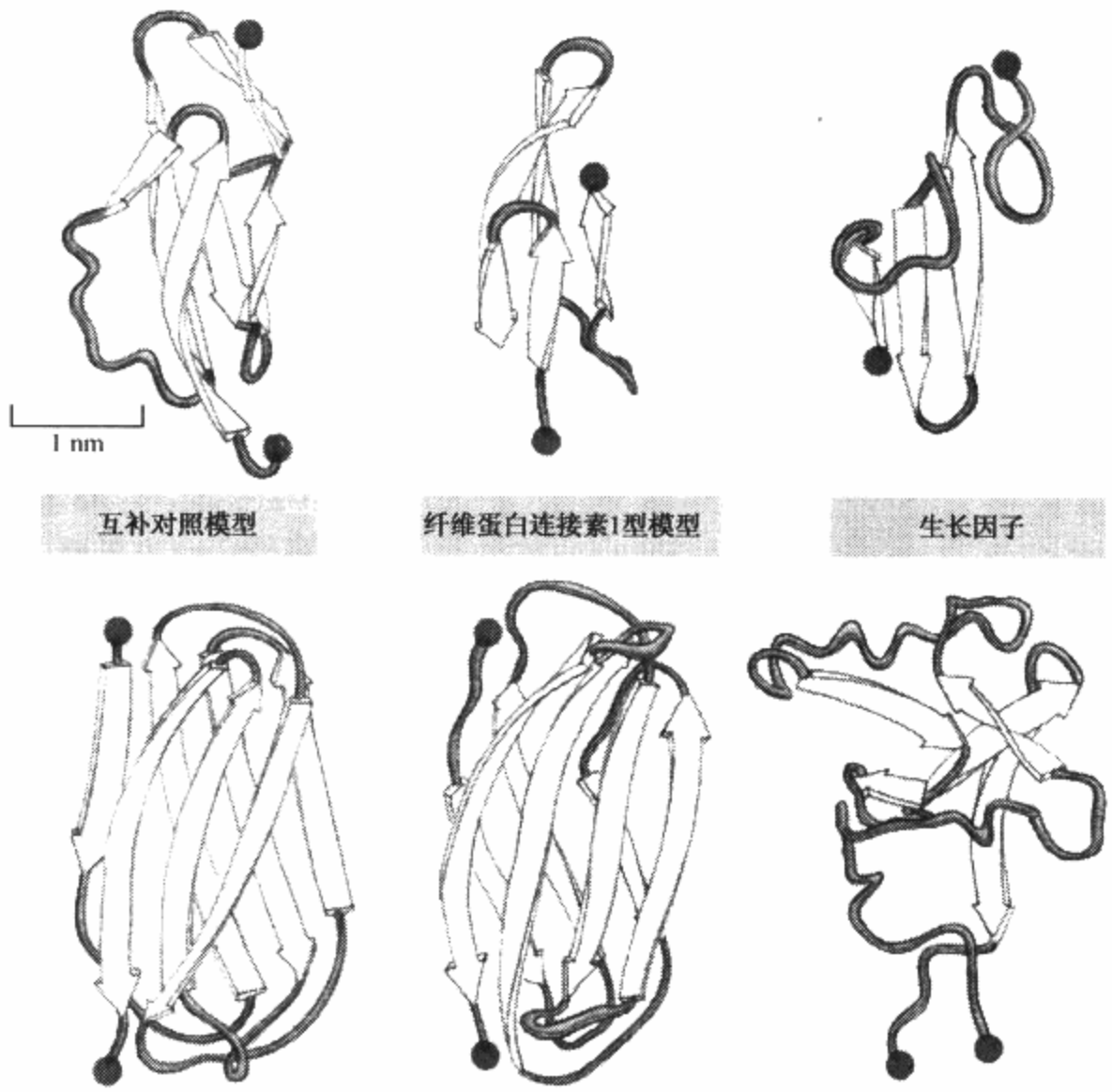


图 3-19　一些蛋白质模块的三维结构

在这些丝带图形中，β折叠以箭头表示，N 端和 C 端以红球表示

人类基因组编码一套复杂的蛋白质，很多仍然是未知

人类基因组测序的结果惊人地表明我们的染色体只包含 30 000～35 000 个基因。根据基因数量，我们比小的芥菜植物——拟南芥的复杂性高不到 1.4 倍，比线虫高不到 2 倍。基因组测序的结果表明脊椎动物几乎所有的蛋白质结构域来自于非脊椎动物，只有 7%的已知的人类的结构域是脊椎动物特异的。

但是人类的每个蛋白质是平均相对比较复杂的。脊椎动物进化中发生的结构域洗牌产生了很多蛋白质结构域新的组合，人类中发现的新的蛋白质结构域的组合是线虫或果蝇的 2 倍。例如，类胰蛋白酶丝氨酸蛋白酶结构域，在人类中至少和 18 种其他类型的蛋白质结构域相关，而在线虫中只发现与 5 种不同的结构域共价相连。人类蛋白质的额外的多样性极大地增加了可能的蛋白质-蛋白质相互作用的范围（图 3-78），但是它如何构成我们人类仍然是不得而知的。

活的有机体的复杂性是让人难以置信的，人类基因组中还有 10 000 多个蛋白质需要鉴定，我们应该清醒地认识到我们对它们的功能了解是非常少的。摆在未来的细胞生物学家面前的挑战是巨大的，他们必须勇敢地来揭开这个神秘的世界。

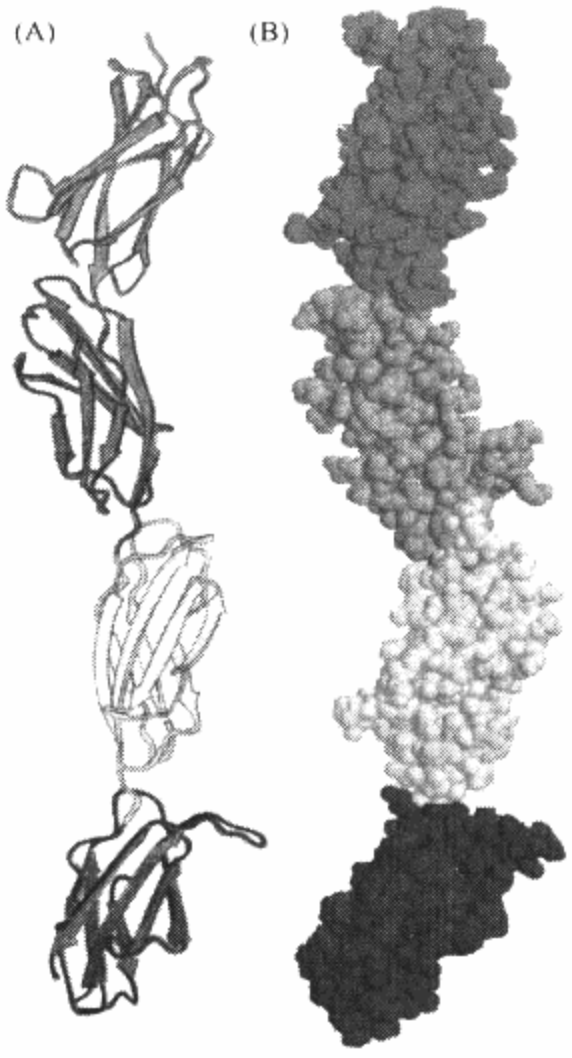


图 3-20　线性蛋白质模块的伸展结构

4（纤连蛋白）3 型模块（图 3-19）来源于胞外基质（纤连蛋白）分子。（A）丝带模型；（B）空间占有模型（摘自 D. J. Leahy, I. Aukhil and H. P. Erickson, *Cell* 84：155～164，1996）

大的蛋白质通常包含不止一条多肽链

使一条多肽链折叠成为一个特殊构象的弱的非共价键同样也能使几个蛋白质相互结合而形成细胞中更大的结构。一个蛋白质表面的任何一个区域都可以和别的分子通过一个称为**结合位点**的非共价键相互作用。一个蛋白质可以包含很多或大或小的针对不同分子的结合位点。如果一个结合位点识别了另一个蛋白质的表面，它们在此结合位点的紧密的结合作用可以使它们形成一个具有精确空间构象的大的蛋白质分子。这种蛋白质的每一条多肽链称作一个**蛋白质亚基**。

最简单的情况是，两个相同的折叠的多肽链依靠两个相同的结合位点，以“头对头”的形式相互结合，形成一个包含两个蛋白质亚基的对称的复合物。Cro 阻遏蛋白——基因调节蛋白，其和 DNA 结合而关闭基因在细菌细胞内的表达（图 3-21）。细胞中，有很多由一条多肽链的多个拷贝形成的其他类型的对称的蛋白质复合物。例如，神经氨酸酶由 4 条相同的蛋白质亚基组成，相互之间以“头尾”排列形成一个紧密的环形（图3-22）。

细胞中还存在很多由两条或多条不同多肽链组成的蛋白质。例如，在红细胞中携带氧气的一种蛋白质——血红蛋白，是研究的很好例子（图 3-23）。它由两条相同的 α 链和两条相同的 β 链对称排列而形成。这样多亚基的蛋白质复合物在细胞中是很常见的，而且它们可以形成很大的分子。图 3-24 显示的是一些已知其精确三维结构的蛋白质，可以比较我们作为模型的相对小的蛋白质与这些大的蛋白质的大小和形状。

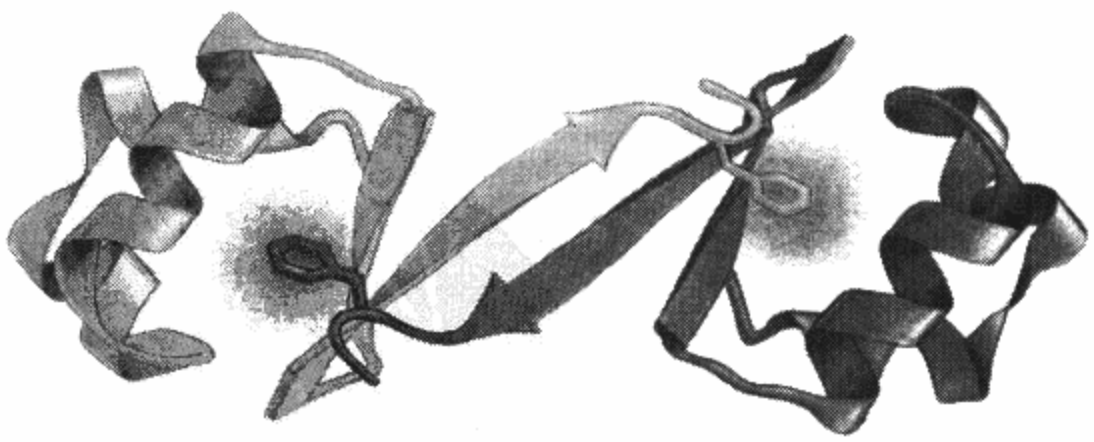


图 3-21　两个相同的蛋白质亚基相互结合形成一个对称的蛋白质四聚体

λ 噬菌体的 Cro 阻遏蛋白结合到 DNA 上关闭病毒基因的表达。其两个相同的亚基头对头结合，通过疏水作用（蓝色）和氢键（黄色区域）结合到一起

图 3-22　一个蛋白质分子包含多个拷贝的蛋白质亚基

神经氨由 4 条相同的多肽链组成一个环状结构。小的示意图表明了相同的相互作用是如何形成稳定的结构的

一些蛋白质形成长的螺旋状纤维

一些蛋白质分子可以组装为缠绕在整个细胞内的纤维。最简单的情况，如果每个蛋白质分子有一个结合到另一个相同分子表面的互补的结合位点，那么就可以形成一个长的蛋白质分子（图 3-25）。例如，肌动蛋白质丝是由很多肌动蛋白质分子形成的长的螺旋状结构分子（图 3-26）。真核细胞中肌动蛋白非常丰富，它构成了细胞骨架中一个主要的纤维系统（讨论见第 16 章）。

为什么螺旋是一个常见的结构呢？据我们所知，生物学结构通常由相似的亚基连接，如氨基酸或蛋白质分子成为长的、重复的链。如果所有的亚基是相同的，链内的相邻的亚基为了最大限度地减小它们之间的自由能，只能调整其相对位置成为唯一的排列方式。因此，每个亚基与相邻的亚基的排列方式基本相同，即第 3 个亚基结合到第 2 个亚基的方式与第 2 个

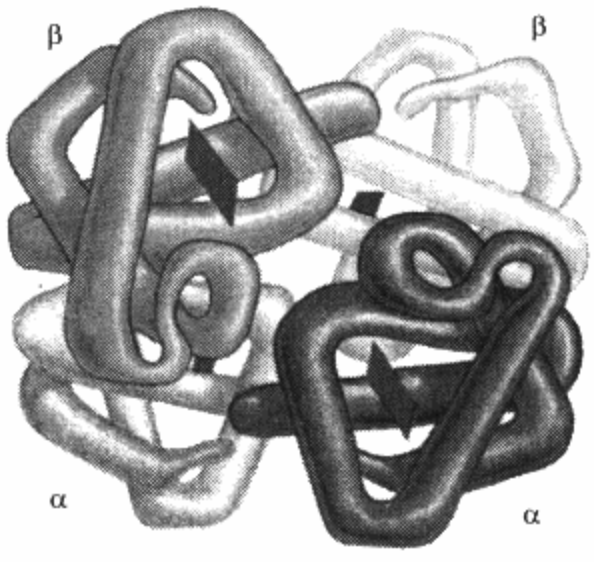


图 3-23　一个蛋白质分子由 2 个不同的亚基形成一个对称的结构

血红蛋白是在红细胞中含量很丰富的一种蛋白质，它包含 2 个 α 亚基和 2 个 β 亚基。这 4 条多聚肽链的每一条都含有一个血红素分子（红色），它是氧分子的结合部位。因此，每分子的血红蛋白携带 4 分子的氧

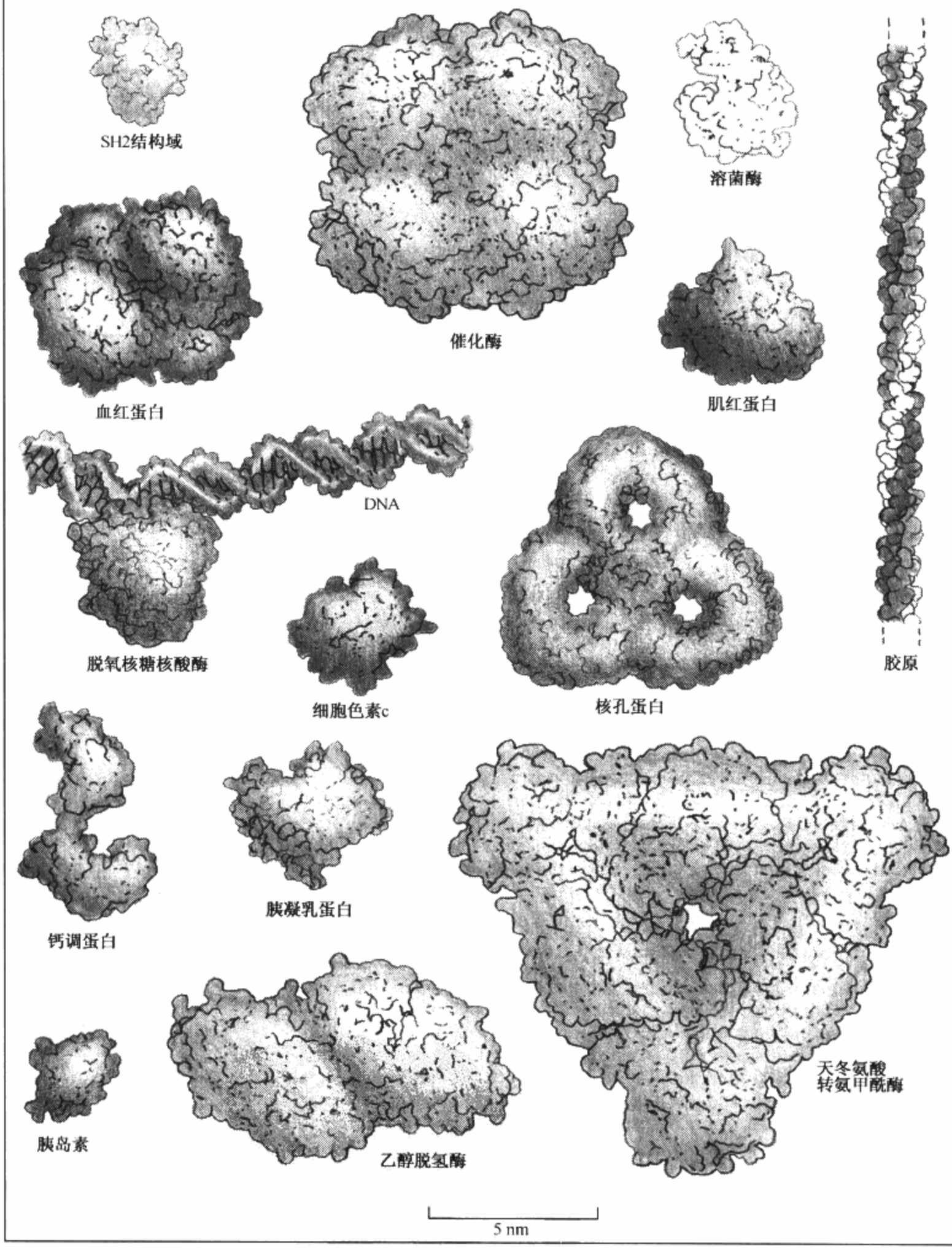


图 3-24　一些蛋白质分子（比例相同）

为了便于比较，DNA 分子结合到蛋白质分子上也有显示。这些空间占有模型显示了一系列的大小和形状。血红蛋白、催化酶、核孔蛋白、乙醇脱氢酶和天冬氨酸转氨甲酰酶都是由多个亚基组成的。SH2 结构域（左上方）在专题 3-2 中有详细说明（After David S. Goodsell, Our Molecular Nature. New York: Springer-Verlag, 1996.）

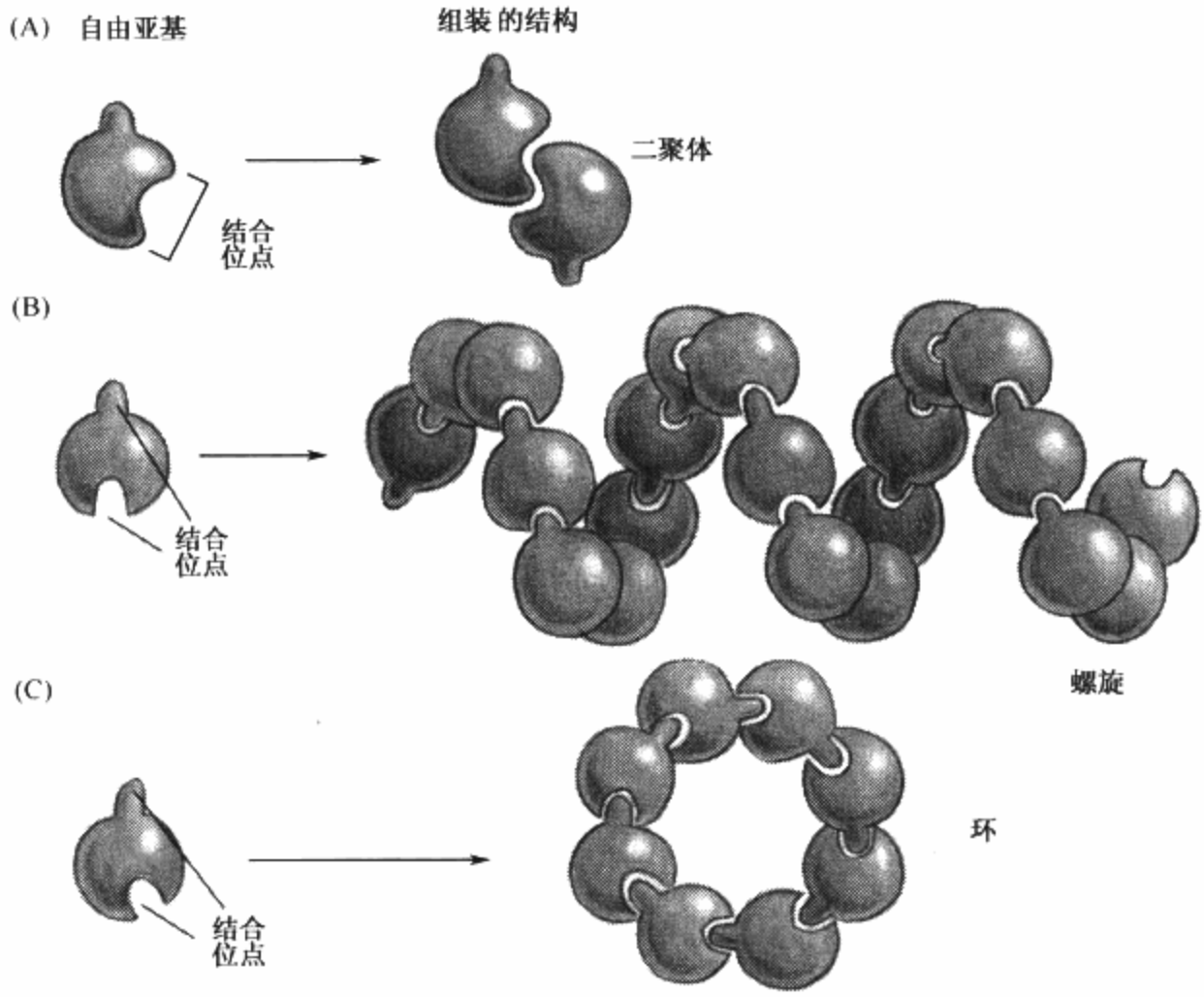


图 3-25　蛋白质组装

（A）只有一个结合位点的蛋白质和另一个相同的蛋白质形成二聚体。（B）有不同结合位点的相同的蛋白质形成长的螺旋状纤维。（C）如果两个结合位点位置接近，蛋白质亚基就形成一个闭合环状结构

亚基结合到第 1 个亚基的方式相同，依此类推。又因为亚基的线性排列是非常少见的，这样就产生了螺旋这种规则的序列——类似于螺旋梯，如图 3-27 所示。由于螺旋梯的扭曲，可以有右手螺旋或左手螺旋［图 3-27（E）］。上下颠倒螺旋不会影响其手性，但是在镜子中其手性会逆转。

生物大分子中螺旋是很常见的，或者由小分子依靠共价键连接（如氨基酸形成 α 螺旋），或者由非共价键连接形成大的蛋白质分子（如肌动蛋白纤维中的肌动蛋白分子）。这并不令人惊讶。螺旋是一个普通的结构，它的产生是很多相似的亚基相邻排列的结果，每 2 个之间的关系是相同的。

一个蛋白质分子可以形成长的、纤维状结构

我们前面讨论的大部分蛋白质都是球状蛋白，多肽链折叠为一个有不规则表面的，紧密的类似于球的形状。酶一般都是球状蛋白：虽然酶包含很多亚基，而且大而复杂，但是大部分都是球形（图 3-24）。相反地，细胞中的另一些蛋白质需要跨越

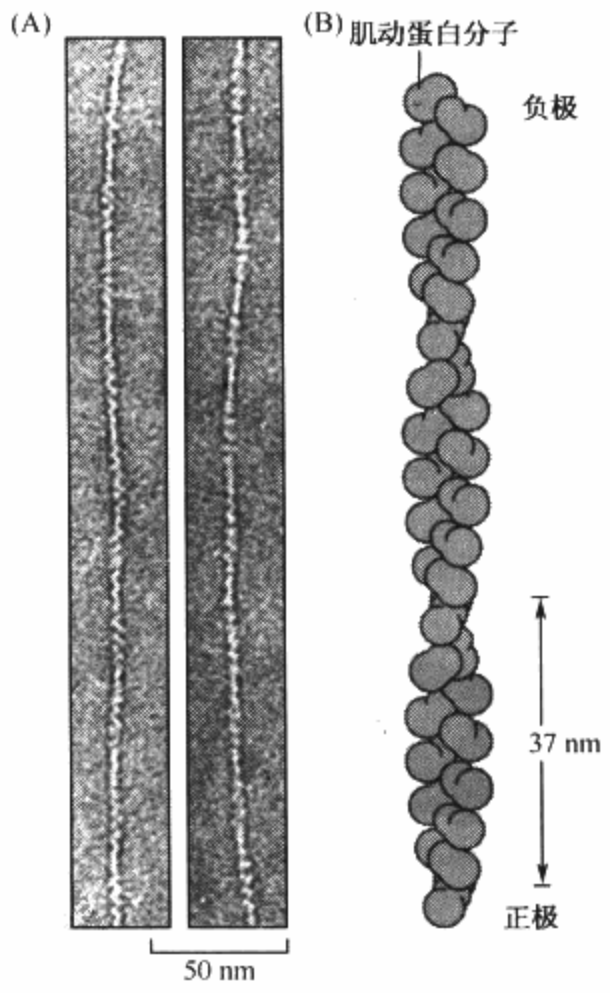


图 3-26　肌动蛋白纤维

（A）透射电子显微镜下的负染的肌动蛋白纤维。（B）肌动蛋白纤维中螺旋排列的肌动蛋白分子（经 Roger Craig 许可）

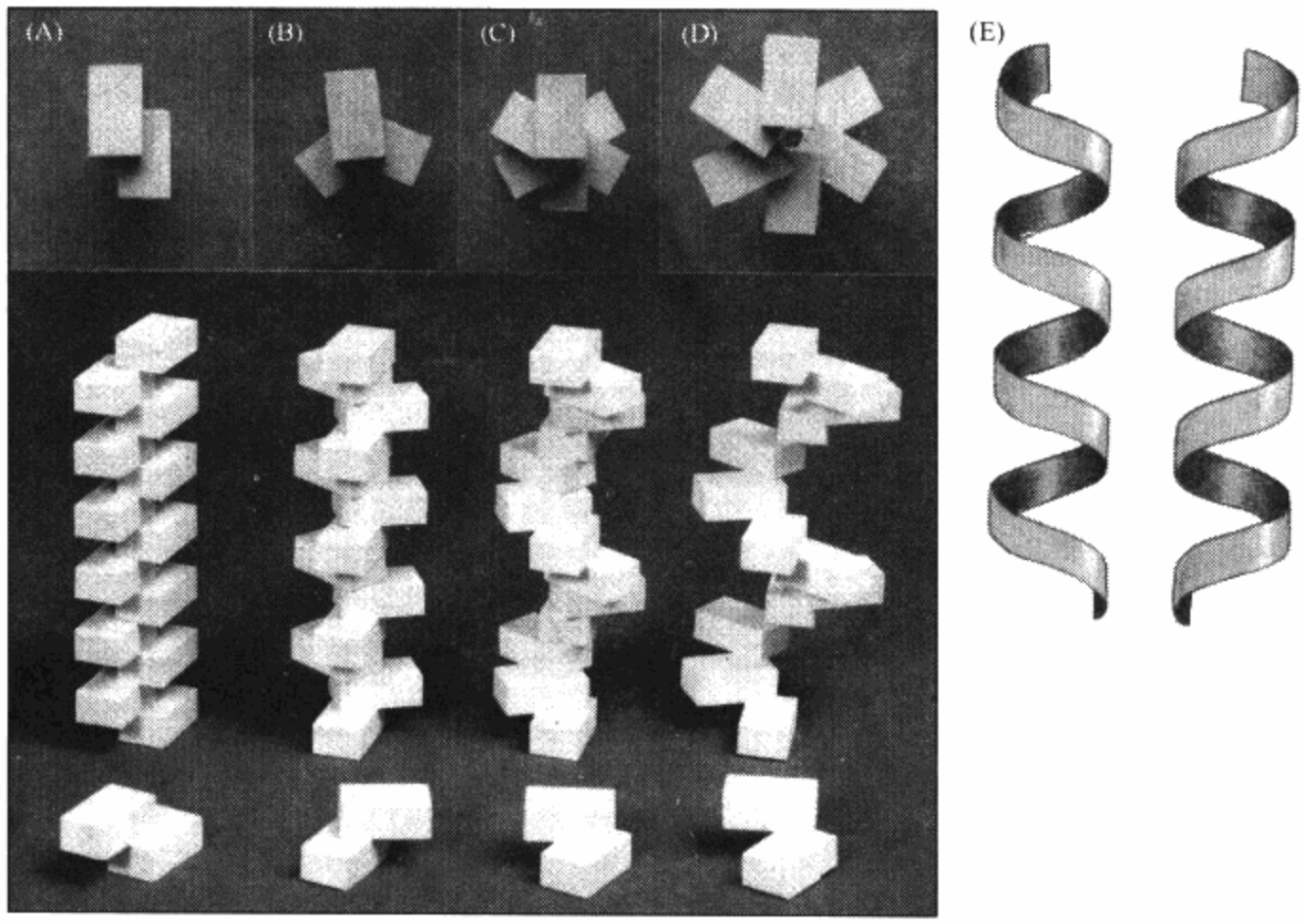


图 3-27　螺旋的一些特性

（A）～（D）当一些亚基以一种规则的方式相互结合则会形成螺旋。底部指出了 2 个亚基的相互作用；其后面即使形成的螺旋。这些螺旋每圈含有 2 个（A）、3 个（B）、6 个（C 和 D）亚基。顶部，直接从螺旋的上方观察亚基的排列方式。注意到（D）图螺旋的宽度大于（C）图，但是它们每一圈的亚基数是相等的。（E）螺旋可以是右手性或左手性的。作为参考，记住标准的金属螺丝钉，当以顺时针方向拧，则表明是右手螺旋。当一个螺旋上下颠倒时，并不改变其手性

很长的距离。这些蛋白质通常形成相对简单的，有长的三维构象，通常叫做纤维状蛋白。

胞内蛋白中纤维状蛋白的一个大的家族包含 α-角蛋白以及其成员。角蛋白纤维非常稳定，是头发、角、指甲等生活周期较长的结构的主要组分。一个 α-角蛋白分子由两个相同的亚基组成，每个亚基的 α 螺旋形成 coiled-coil 结构（图3-11）。在 coiled-coil 结构的末端有球状结构域，这个结构域上正好有结合位点。这种类型的蛋白质可以组装为绳状中间纤维——细胞骨架中一个非常重要的组分，它构成胞内的支架结构（图 16-16）。

纤维状蛋白在细胞外含量很丰富，它们是辅助由细胞团形成组织的胶状细胞外基质的主要成分。细胞外基质蛋白是由细胞合成后分泌到胞外的，它们通常形成片状或长的纤丝。胶原蛋白是动物组织中含量最丰富的蛋白质。一个胶原分子是由三条长的多肽链组成，每条链中每三个氨基酸的第三位都是甘氨酸。这种规则结构使多肽链相互缠绕形成三股螺旋［图 3-28（A）］。很多胶原分子首尾相接，并行排列组成极其坚固的胶原纤维，形成具有抗张强度的结缔组织，讨论见第 19 章。

弹性蛋白是细胞外基质中另一个组分。弹性蛋白分子由相对松散和无规则结构的多肽链共价交联形成一个弹性蛋白网络：不像大多数蛋白质，弹性蛋白没有一个非常特异或稳定的结构，但是它们可以从一个构象变为另一个相反的构象，如图 3-28（B）所示。由于这些弹性纤维，可以使皮肤和其他组织（动脉、肺）自由地收缩。

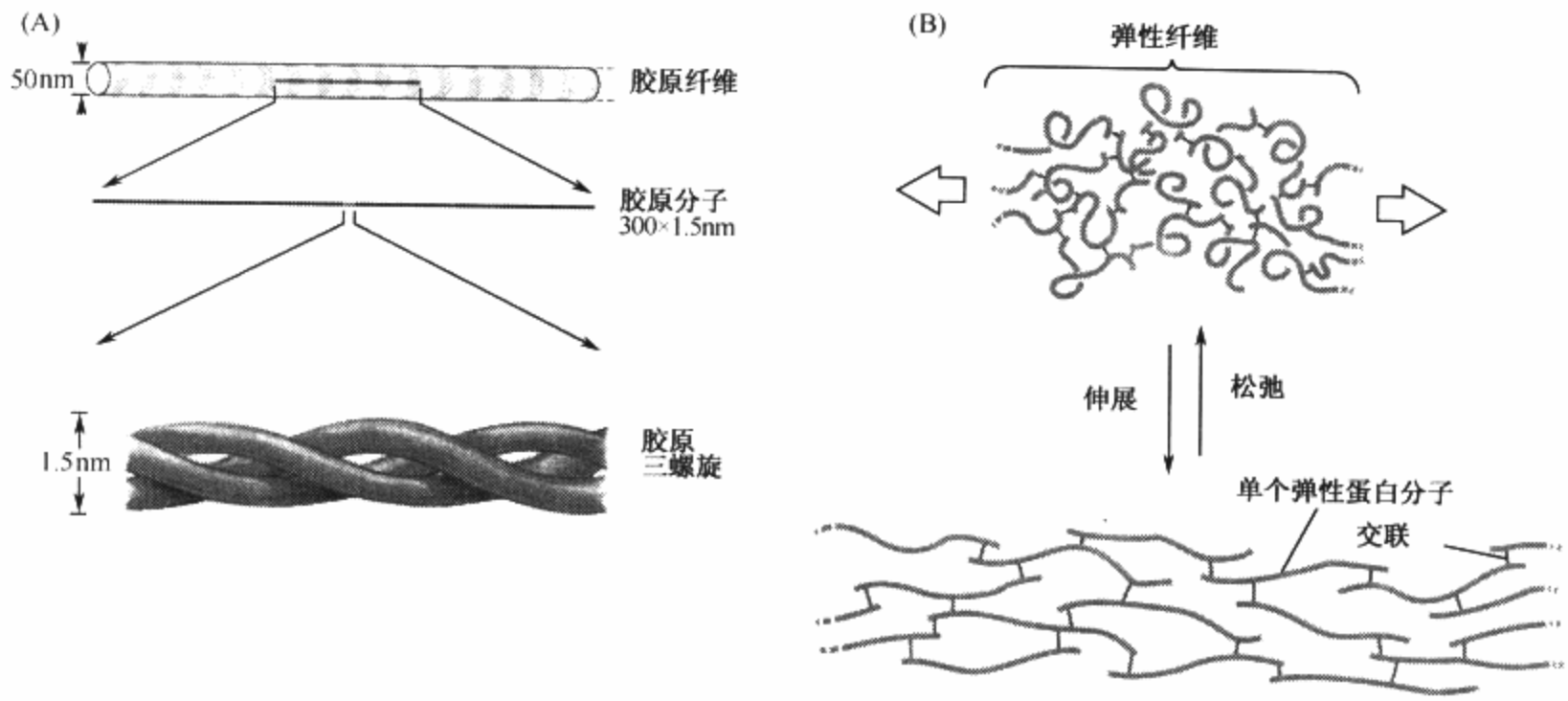


图 3-28　胶原蛋白和弹性纤维

（A）胶原蛋白是由三条伸展的肽链相互缠绕而形成的三股螺旋（底部）。许多这种棒状的胶原分子在胞外空间相互交联形成不可伸展的胶原纤丝（顶部），有钢铁般的强度。胶原纤丝上的条纹是由于胶原分子的有规律的重复排列形成的。（B）弹性蛋白多肽链交联形成弹性纤维网络。当纤维受到外力时，每个弹性蛋白分子展开成一个伸展的构象。当拉力消失时，其又自发地缩回

共价交联稳定胞外蛋白

许多蛋白质分子或者与细胞质膜外表面相连，或者被分泌到胞外作为细胞外基质的一部分。这些蛋白质都直接暴露在胞外环境中。为了维持它们的结构，这些蛋白质的多肽链通常需要共价交联来稳定其结构。这些交联可以是连接同种蛋白质的两个氨基酸，也可以是多亚基蛋白质连接不同的多肽链。最常见的蛋白质共价交联是共价的硫-硫键。当蛋白准备输出到细胞外时就会形成这种二硫键（也叫做 S—S 键）。如第 12 章所述，在内质网中，二硫键由一种酶催化合成，通过连接蛋白质上相邻的半胱氨酸残基的两个—SH 基团而形成（图 3-29）。二硫键并没有改变蛋白质的构象，反而作为一个原子钉可以帮助形成最合适的构象。例如，溶菌酶——一个存在于眼泪中可以降解细菌细胞壁的酶，由于这种交联可以长时间维持其抗菌活性。

在细胞质中二硫键一般很难形成，高浓度的还原剂可以将 S—S 还原成为巯基(—SH)。显然，在细胞内相对温和的环境中，蛋白质不需要这种强化作用。

蛋白质分子通常作为大的结构分子的亚基

蛋白质分子自身相互连接可以形成细胞内的大结构分子——超分子结构，如酶复合物、核糖体、蛋白质纤维、病毒和膜。这些物质不是由简单的大的分子共价连接而成。它们是由那些作为最终结构亚基的独立的分子依靠非共价作用组装而成的。

利用小的亚基形成大的结构物有很多优点：

1. 利用一个或多个重复的小亚基组装大的结构只需要较少的遗传信息；

2. 因为亚基是通过能量较低的键连接起来的，它们的装配和解聚是一个可逆的过程，很容易控制；

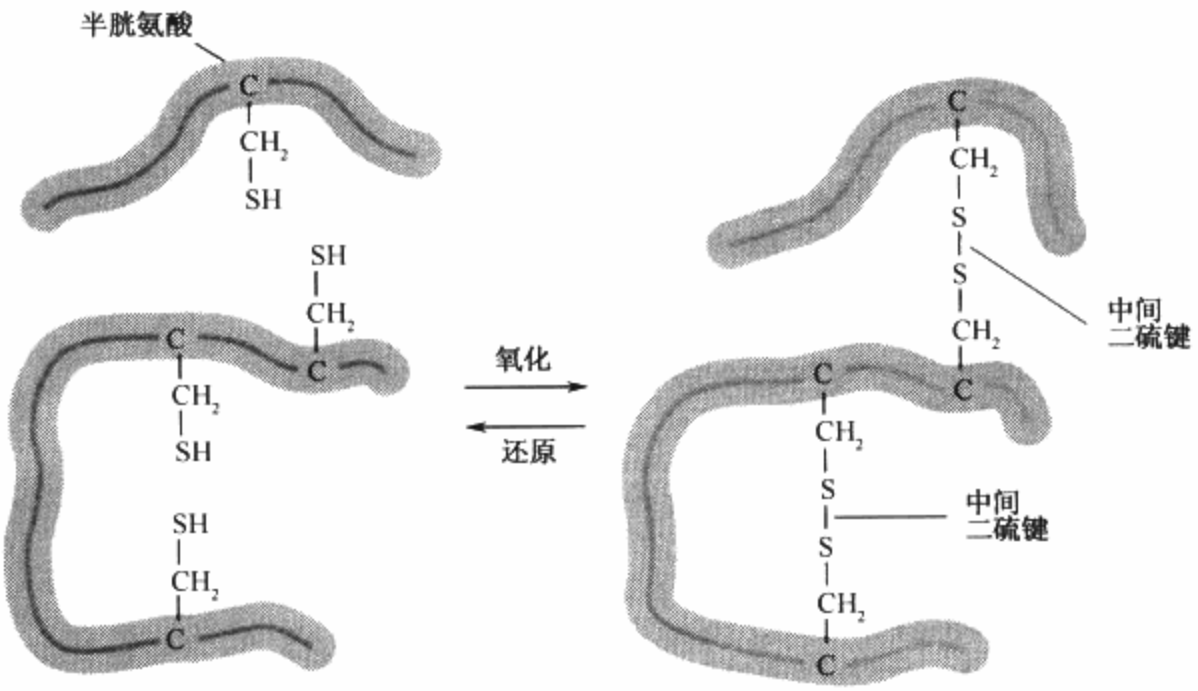


图 3-29　二硫键

该图表明了两个相邻的半胱氨酸残基怎么形成共价二硫键。如图所示，一条多肽链的两个部分或者两条不同的多肽链之间都可以交联。因为破坏共价键所需要的能量远远高于非共价键所需的（表 2-2），所以二硫键对于稳定蛋白质具有重要的作用

3. 可以避免结构合成过程中的错误，因为在装配的过程中正确的机制可以排除形成的畸形结构。

一些蛋白质亚基可以以六边形排列，装配为平的片状结构。特殊的膜蛋白在脂质双分子层中通常以这种方式排列。如果每个亚基的几何结构发生很小的变化，就会从片状变为管状结构（图 3-30），如果变化再大一点，则会形成一个中空的球形。管状和球状蛋白可以结合特异的 RNA 或 DNA 分子而形成病毒的核心蛋白质。

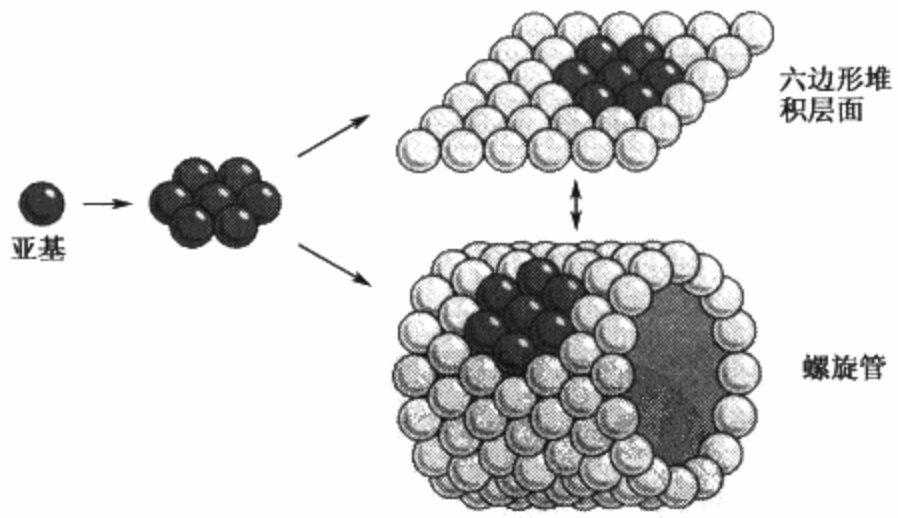


图 3-30　单一蛋白质亚基组装需要多个蛋白质之间的相互作用

六边形堆积形成的球状蛋白质亚基可以形成一个片状结构或者一个管状结构

装配的紧密的结构，如环状、管状、球状等，具有额外的稳定性，因为它增加了蛋白质亚基之间的化学键。而且，这样的结构是依赖于亚基之间的相互作用，每个亚基的小的变化可以装配或解聚这个大的结构物。在简单病毒的核心或*壳体*蛋白质中，这是很常见的，它们通常形成一个中空的球状结构（图 3-31）。壳体蛋白质由数百个相同的蛋白亚基组成，可以保护病毒的核酸（图 3-32）。这种外壳蛋白质必须有特别的适应性结构：它不仅需要多种不同的

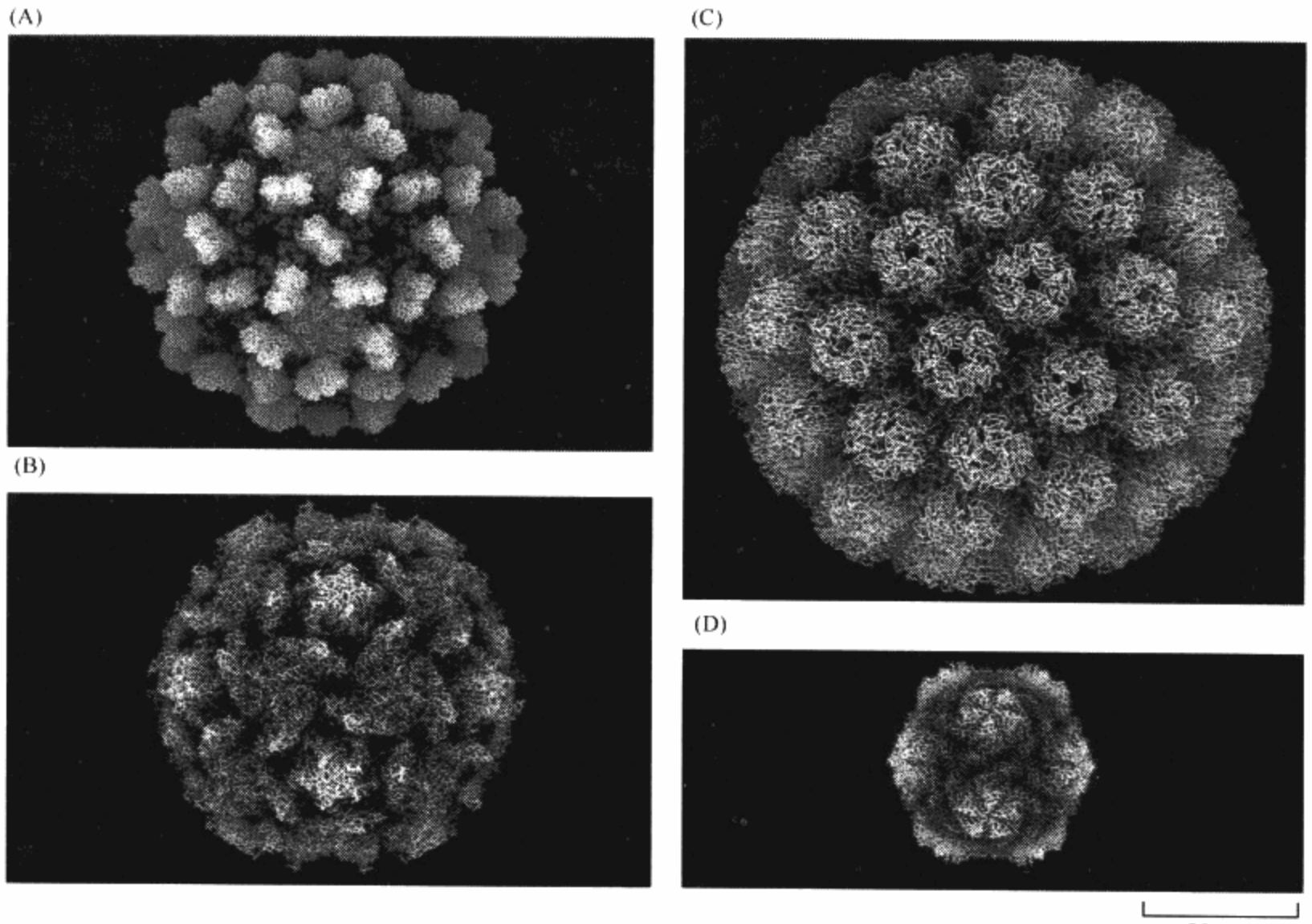


图 3-31　一些病毒蛋白的壳体（比例相同）

（A）番茄丛矮病毒；（B）脊髓灰质炎病毒；（C）猴病毒 40（SV40）；（D）卫星烟草坏死病毒。这些壳体的结构由 X 射线晶体衍射解出，并且都是在原子水平上的分辨率（经 Robert Grant，Stephan Crainic 和 James M. Hogle 许可）

作用来形成球体，而且一旦病毒成功进入细胞，它必须改变这种排列来释放病毒核酸起始其在细胞内的复制。

细胞内的很多结构是自组装的

生物大分子的组装信息是储存在亚基中的，因为纯化的亚基可以在合适的条件下自发组装成其最终的结构。第一个被发现的可以从其亚基自组装为大分子集合体的是烟草花叶病毒（TMV）。此病毒是一个长的棒状结构，形成的圆柱形蛋白质围绕在螺旋状 RNA 外（图 3-33）。如果单独分离出 RNA 和病毒亚基，并将它们混合在溶液中，它们可以组装为有活力的病毒颗粒。这个装配过程很复杂，包括蛋白质形成一个双环的结构作为中间介导物，加到延伸的病毒核心上。

另一个可以自组装的大分子是细菌的核糖体。这个结构包含 55 种不同的蛋白质分子和 3 种不同的 rRNA。如果把每个独立的组分置于试管中，在合适的条件下，它们可以自发形成其原来的结构。更重要的是，这样重组装的核糖体仍然具有合成蛋白质的能力。可以想像，核糖体的重装配也有特殊的方式：当某些蛋白质结合到 RNA 上后，其余的蛋白质可以识别这个复合物，依次类推直到结构完成。

现在我们仍不清楚这种精确的自组装是如何调控的。例如，细胞中的很多结构，似乎有

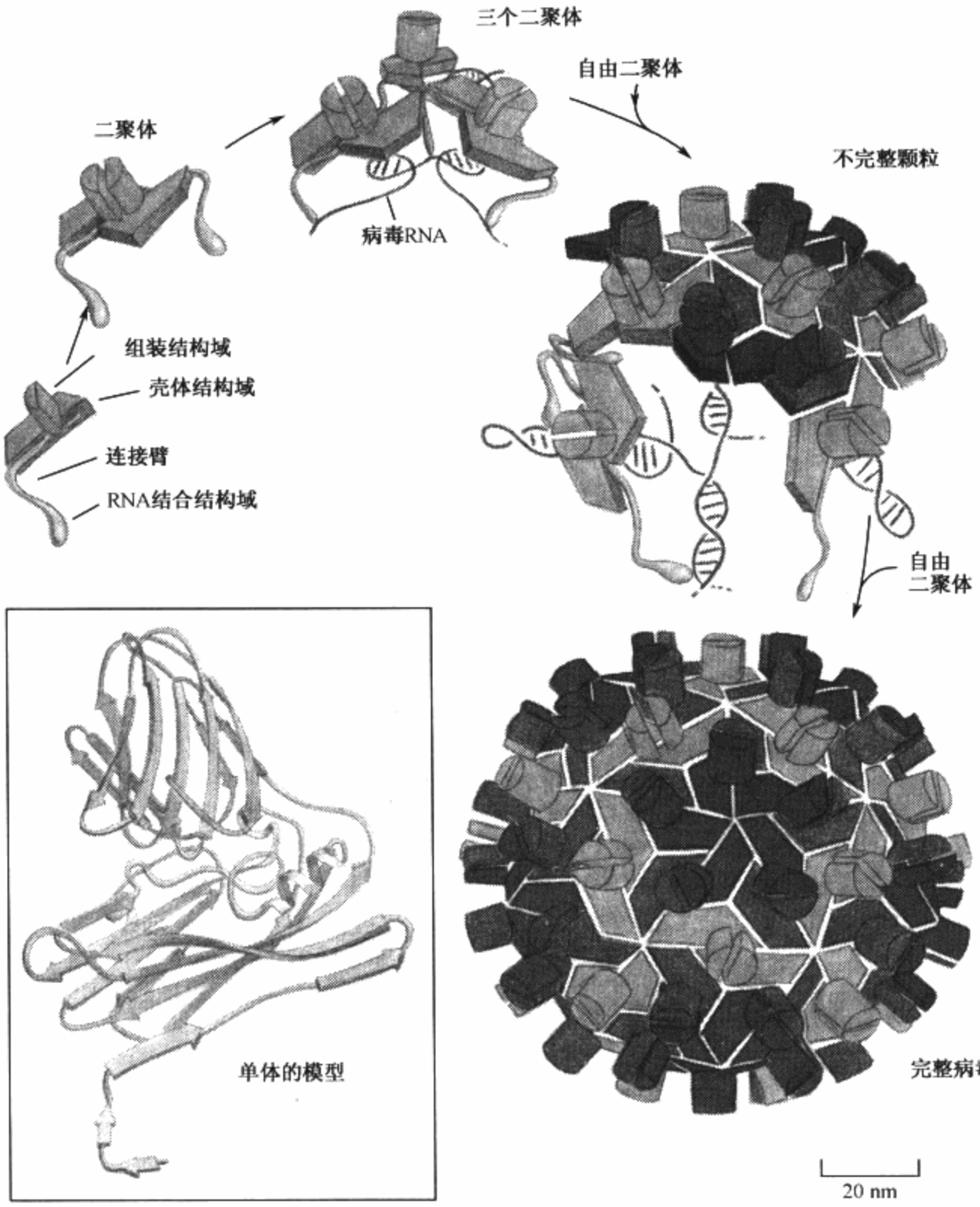


图 3-32　球形病毒的结构

很多病毒是由许多相同的亚基堆积形成的球形外壳（壳体）包裹着 RNA 或 DNA 病毒基因组（图 3-31）。由于空间位阻的原因，不多于 60 个相同的亚基可以以一种精确的对称的方式堆积。但是如果可以允许轻微的不规则，更多的亚基也可以形成更大的壳体。这里以番茄丛矮病毒为例，它是一个球状病毒，直径 33nm，由 180 个相同的每一个包含 386 个氨基酸的壳体蛋白质加上一个 4500 个核苷酸的 RNA 基因组组成。为了形成这样大的一个壳体，蛋白质必须可以适应三种不同的环境，在这里每一种在病毒颗粒中以不同的颜色表示。图中显示了假设的组装方式；精确的三维结构由 X 射线衍射确定（经 Steve Harrison 许可）

一个精确的长度，这一长度比组成它们的生物分子组分要长很多。这种长度的决定是如何实现的仍然是一个谜。图 3-34 揭示了三种可能的机制。最简单的情况，一个长的核心蛋白质或其他大分子提供一个脚手架，决定最终的装配程度。这个机制决定了 TMV 颗粒的长度，而 RNA 链作为主要的核心。类似地，核心蛋白质决定了肌肉中细肌丝的长度，同样也决定了细菌病毒的尾部长度（图 3-35）。

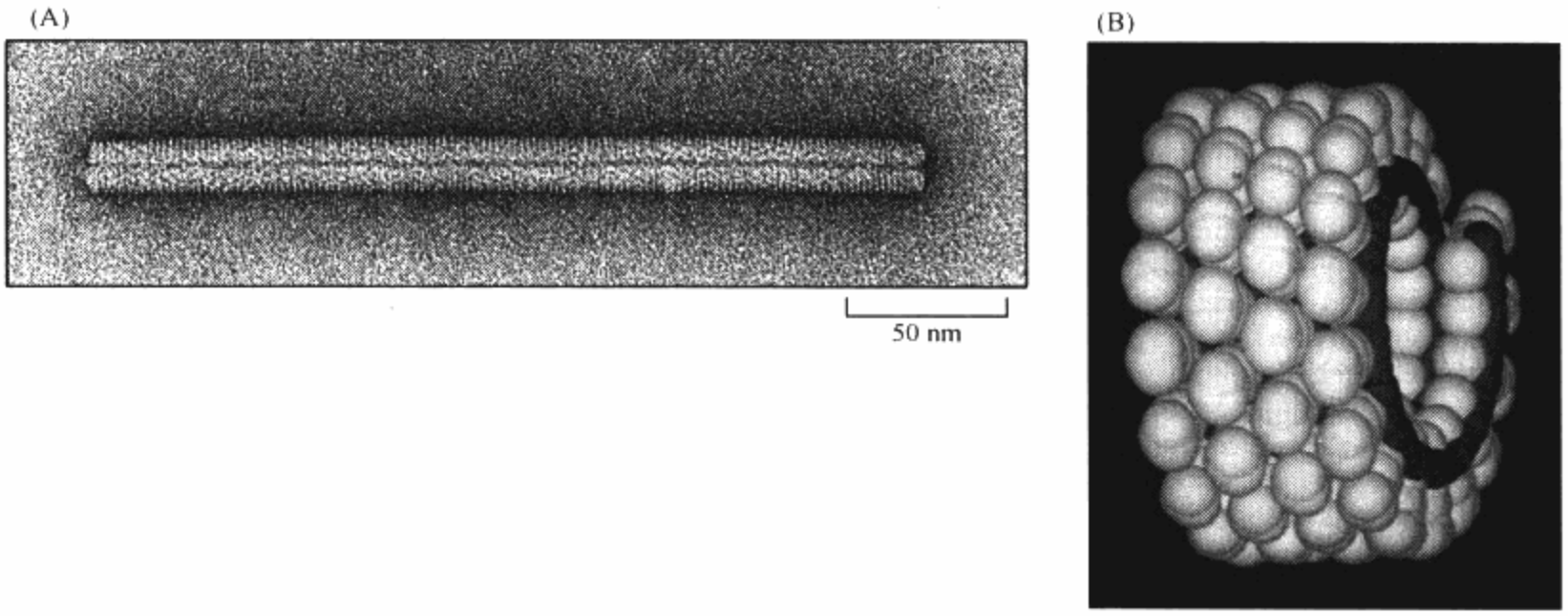


图 3-33　烟草花叶病毒（TMV）的结构

（A）病毒颗粒的电子显微镜图，由相同亚基组成的圆柱状蛋白质外壳包裹着一条长的 RNA 分子。（B）模式图说明 TMV 的部分结构。由 6000 个核苷酸组成的单链 RNA 分子被包装在由 2130 个 158 个氨基酸长的外壳蛋白质形成的螺旋状外壳中。感染性的病毒颗粒可以在含有纯化的 RNA 和蛋白质的试管中自组装[（A）经 Robley Williams 许可；（B）经 Richard J. Feldmann 许可]

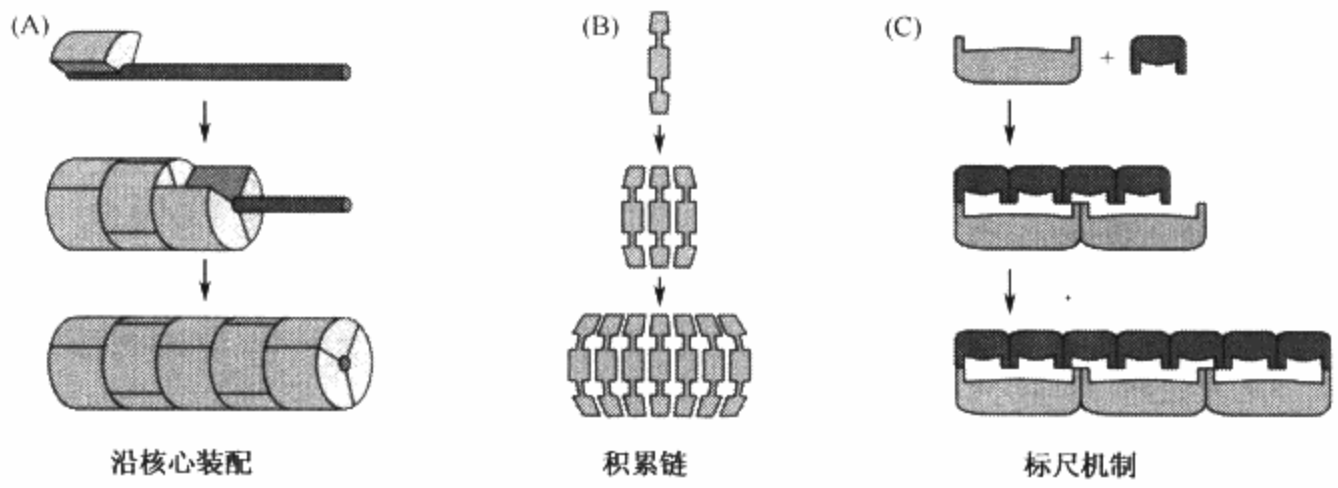


图 3-34　决定大蛋白质装配长度的三种机制

（A）沿着伸长的核心蛋白质或者其他大分子的共组装作为测量的装置。（B）装配的终止是由于当多余的亚基装配时，张力会在多聚结构上积累，超过一定长度之后，再加上一个亚基需要非常高的能量。（C）组装的标尺，两套长度不同的棒状分子形成一个交错的复合物，复合物不断延伸直到末端精确的匹配。其名字来源于在机器制造中一个原理相同的测量仪器

复杂生物结构的形成需要辅助因子的帮助

并非胞内的所有由共价键连接的结构都可以自组装。例如，线粒体、纤毛或肌细胞中的肌原纤维，在含有其相应组分的溶液中不能自发装配。在这些情况下，装配的部分信息是由特殊的酶以及细胞内的其余蛋白质提供，它们行使模板功能，引导结构组装，但是并不出现在最终成熟的结构中。

即使是相对简单的结构也缺少某些自组装必要的因素。某些细菌病毒的形成，如头部，它仅仅由多个相同的单一的蛋白质亚基组成，但是其装配需要到第二种蛋白质形成的支架上组装。因为在成熟的病毒颗粒中第二种蛋白质是不存在的，因此头部结构一旦被分开它就不

可能自发重装配。还有其他例子，在正常的装配过程中蛋白质的切割是一个重要的、不可逆的过程。这种情况在一些小的蛋白质复合物中也是存在的，如胶原结构蛋白和胰岛素（图3-36）。从这些相对简单的例子中，我们可以看到像线粒体或纤毛这样复杂结构的组装涉及大量其他细胞组分的时间和空间的调控。

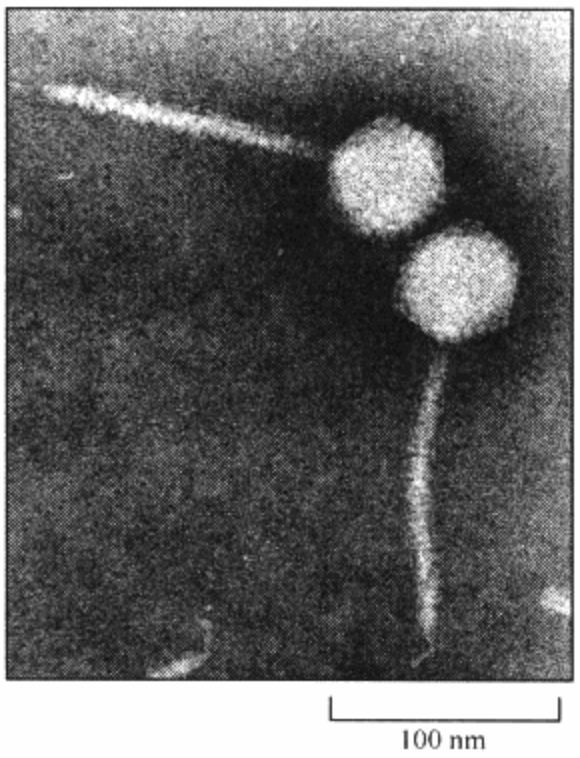


图 3-35　λ噬菌体的电镜图

当病毒尾部的末端结合到一个细菌细胞表面的特异蛋白质上后，被紧密包装在头部的DNA通过尾部被注射到细胞中。尾部的长度是非常精确的，其机制见图3-34（A）

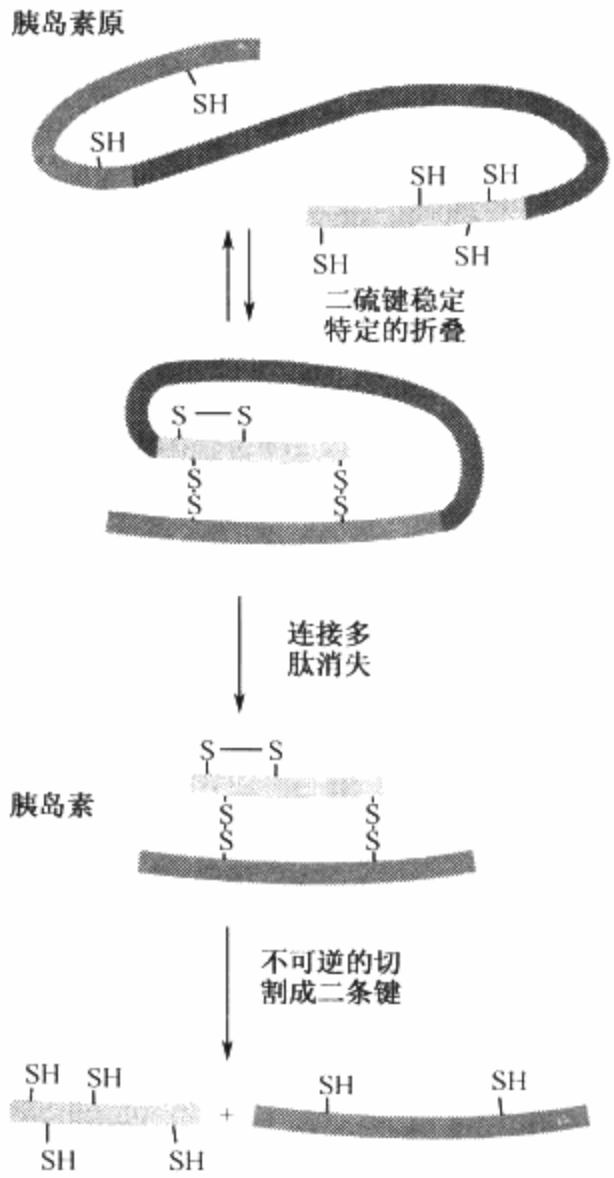


图 3-36　胰岛素组装中的蛋白质切割

多肽激素胰岛素如果其二硫键被破坏是不能自发重新高效装配的。其合成初期是一个大蛋白质（胰岛素原），然后当蛋白质折叠成一个特殊构象时发生蛋白质切割产生胰岛素。当胰岛素原变性时，两条多肽链分开，多肽链切割的部分信息被丢失，蛋白质不能自发形成其正确的构象

小结

蛋白质分子的三维空间构象由其氨基酸序列决定。多肽链不同区域之间的非共价相互作用稳定高度折叠的结构。疏水性氨基酸易于包埋在分子内部，而相邻肽键局部的氢键相互作用，可以形成α螺旋和β折叠。

球状区域，通常称为结构域，是蛋白质的折叠单位；这样的结构域通常包含40～350个氨基酸。典型的小蛋白质通常只包括一个结构域，而大的蛋白质通常由几个结构域，依靠多肽链短的结构连接而成。随着蛋白质的进化，结构域可以发生修饰或与不同结构域之间的重组形成新的蛋白质。参与形成大的蛋白质的结构域也可以叫做模块。目前，在已知的10 000多个蛋白质结构中，已经鉴定出1000种不同的结构域折叠方式。

维持蛋白质折叠的非共价作用也可以将蛋白质组装成更大的结构物。有和自身相互结合

位点的蛋白质可以形成二聚体，闭环状、空壳状和多聚螺旋状结构。虽然蛋白质和核酸的混合物可以在试管中自发装配，但是某些大的生物大分子组装过程涉及不可逆的步骤。不是所有的细胞内大结构分子在它们解聚为其相应的组分之后都可以自发装配。

蛋白质的功能

每种蛋白质特异的氨基酸序列决定其三维构象或结构。但是蛋白质不是坚硬的块状材料，它们有精确的运动部分，其机械运动与其化学功能相偶联。这种化学性质和运动相偶联使得蛋白质在活细胞动态环境中有特殊的能力。

这一部分，我们会讲述蛋白质是如何结合其他特定的分子，以及这种结合与其功能的关系。蛋白质和其他分子的结合能力使得蛋白质可以作为催化剂、信号受体、开关、分子马达或小泵等。这一章所举的例子显示了蛋白质的全部生物学功能。但是在本书中有些蛋白质的特殊功能是基于相似的原理。

所有的蛋白质都可以结合其他分子

蛋白质的生物学性质是依赖于其物理上与其他分子的相互作用。例如，抗体结合到病毒或细菌上来减少其破坏作用，己糖激酶结合葡萄糖和 ATP 来催化它们之间的反应，肌动蛋白分子相互结合来组装肌动蛋白纤维等。实际上，所有的蛋白质都可以黏附或结合其他的分子。有些情况，结合非常紧密；有些情况，结合是很弱的，而且作用时间短。但是这些结合都是非常特异的，所以在每个蛋白质分子遇上的成千上万种分子中，蛋白质通常只结合一种或少数几种。蛋白质结合的底物，不论是离子、小分子或大分子都是此蛋白质的**配体**（来源于拉丁语 ligare，意思是结合）。

蛋白质结合的选择性以及对其配体的高亲和性依赖于弱的、非共价键（氢键、离子键）的网络的形成和范德华力，加上疏水相互作用（专题 2-3）。因为每个单独的键是很弱的，所以有效的结合作用需要大量弱的作用力同时形成。只有当配体分子的表面轮廓与蛋白质很相似时，就像一只手在手套中，这种情况才成立（图 3-37）。

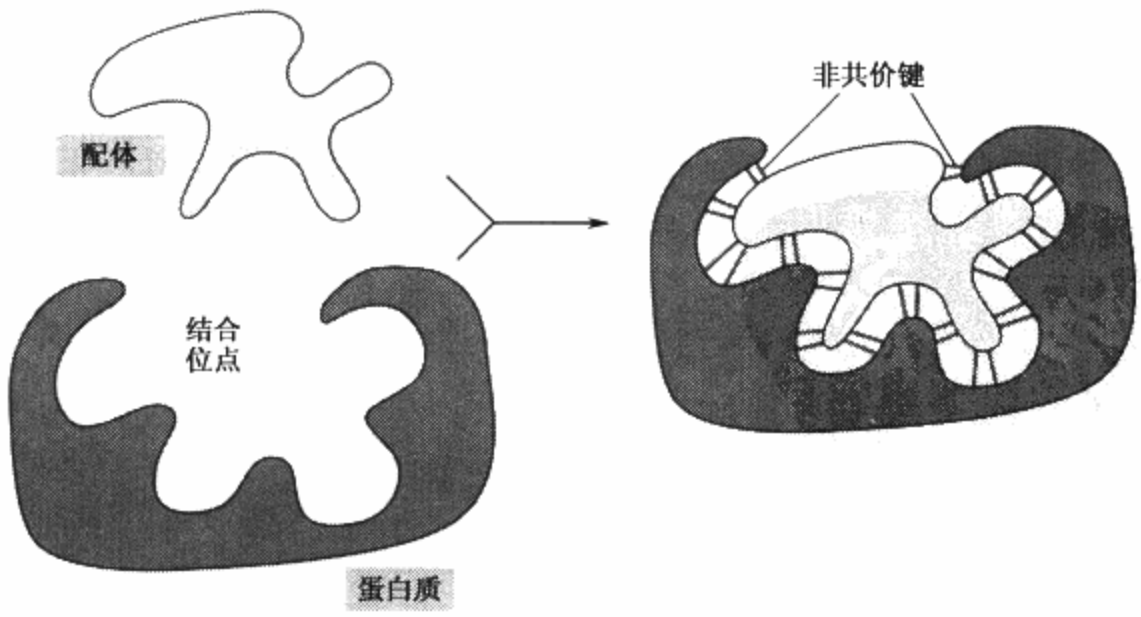


图 3-37　蛋白质选择性地结合其他分子

很多弱的相互作用可以用来保证蛋白质和第二个分子的紧密结合，这第二个分子也叫做蛋白质的配体。配体必须和蛋白质的结合位点精确配对，就像手和手套一样，所以蛋白质和其配体之间可以形成大量的非共价键

蛋白质结合配体的区域叫做配体的结合位点，它通常是由氨基酸的特殊排列形成的蛋白质表面的洞穴。这些氨基酸分布在蛋白质折叠时多肽链的不同区域（图 3-38）。蛋白质表面的不同区域有结合不同配体的不同结合位点，可以调节蛋白质活性。蛋白质的其余部分作为把手将蛋白质置于胞内的特定位置，如前面讲过的 SH2 结构域，它驱使包含它的蛋白质定位于细胞质膜来应答特殊的信号。

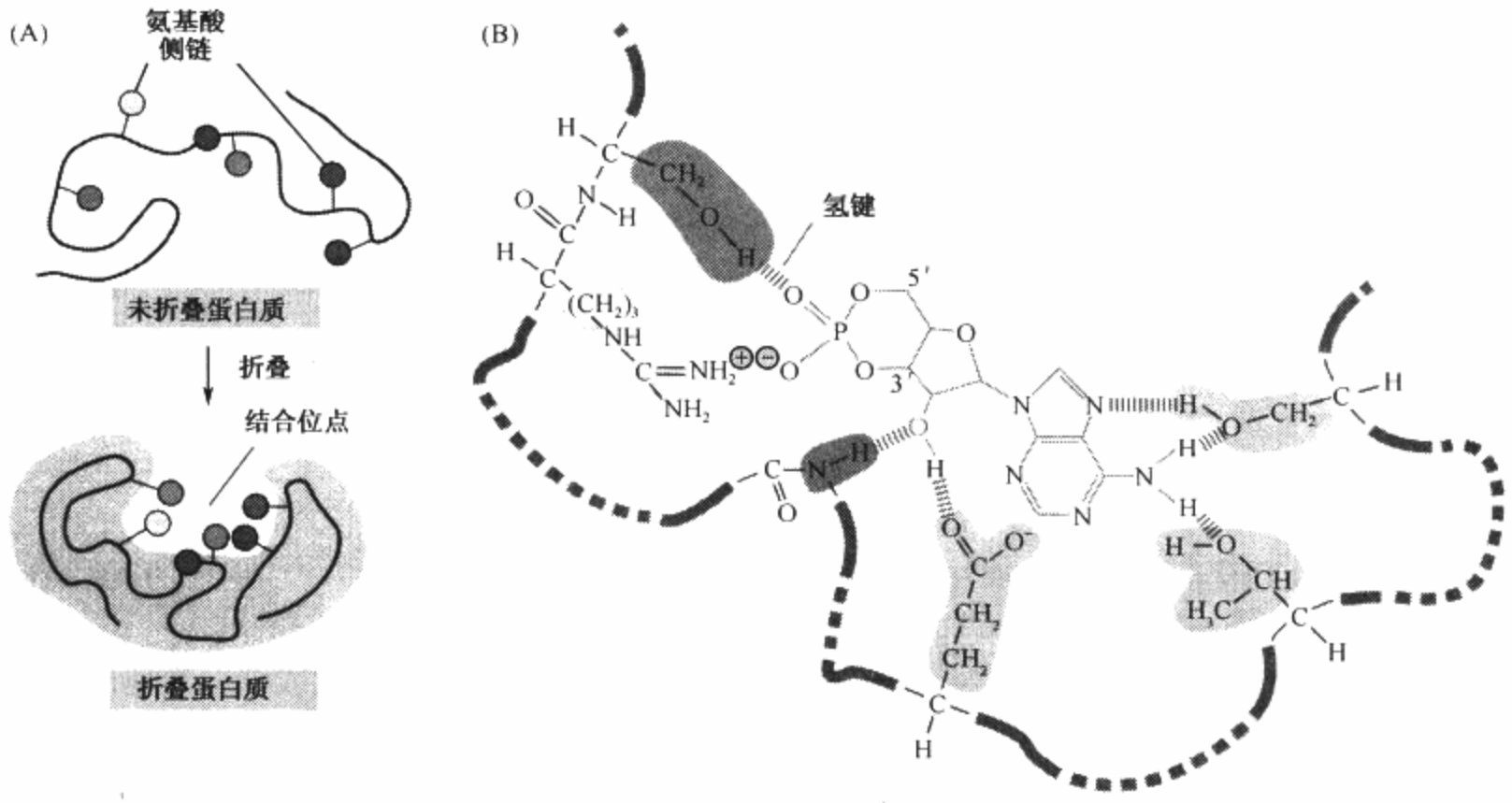


图 3-38　蛋白质的结合位点

（A）多肽链的折叠很容易在蛋白质表面产生一个缝隙或者缺口。这个缝隙中的氨基酸链在和特定的配体结合时会形成非共价键。（B）结合位点的详细图示。蛋白质和其相应配体之间的氢键和离子键（在该例子中，环腺苷酸是结合的配体）

虽然包埋在蛋白质内部的原子与配体没有直接的联系，但是它作为重要的脚手架赋予蛋白质表面轮廓和化学性质。内部氨基酸的细微变化就可能导致其三维结构的变化，而破坏其表面结合位点。

蛋白质构象的细节决定了其化学性质

蛋白质具有重要的化学性质，其分子表面邻近的化学基团可以与之相互作用而增强氨基酸侧链的化学反应活性。这些相互作用分为两类：

第一类，多肽链相邻的部分会以一种作用方式限制水分子进入配体结合位点。因为水分子易于形成氢键，它们可以与配体竞争蛋白质表面的位点。如果排除水分子，蛋白质与其配体之间紧密的氢键（和离子键）相互作用会增强。因为水分子存在于一个大的氢键网络，最初很难想像有一个方式来从蛋白质表面排除类似水分子这样的小分子，而不影响其配体结合活性（专题 2-2）。实际上，配体结合位点可以处于很干燥的环境，因为如果水分子一定要进入到蛋白质表面的裂缝中，就必须从其氢键网络中分离出来，但是这在能量上是很不经济的。

第二类，相邻极性氨基酸的成簇排列可以改变其活性。例如，一些带负电的氨基酸不采用在蛋白质折叠过程中相互排斥的方式，而硬性结合到一起，那么它们结合带正电离子的亲和力就会提高。另外，当氨基酸依靠氢键相互作用时，一些正常情况下不具有活性的侧链基团（如图 3-39 中丝氨酸的—CH_2OH）可以变为有活性，这样促使它们进入到形成或者破坏

共价键的反应中来。

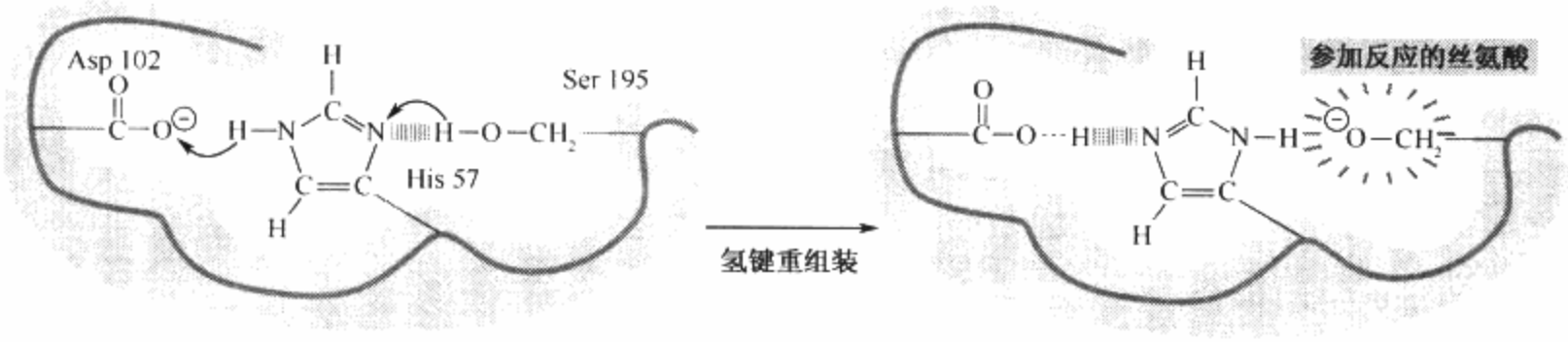


图 3-39　酶活性中心的有明显活性的氨基酸

这个例子是“催化三元组”，存在于胰凝乳蛋白酶、弹性蛋白酶以及其他丝氨酸蛋白酶中（图 3-14）。天冬氨酸侧链（Asp 102）诱导组氨酸（His 57）夺取 195 位的丝氨酸上的质子。激活丝氨酸和酶底物形成共价键，水解肽键

因此每个蛋白质分子的表面都有一个独特的化学反应活性，它不仅依赖于氨基酸的特异性，而且依赖于其相对于别的氨基酸的位置。因为这个原因，即使同一蛋白质的两个差异很小的构象可能在其化学性质上有很大的差别。

蛋白质家族成员序列比对可以发现重要的配体结合位点

如前所述，根据蛋白质的结构域可以划分蛋白质家族，表示它们从一个共同的祖先进化而来，基因组序列表明大多数蛋白质包含一个或多个结构域。同一个家族成员的三维结构是非常相似的。例如，即使氨基酸的相似性只有 25%，结构域中主链原子在 0.2nm（2Å）形成相同的蛋白质折叠方式。

根据这些事实，一种“进化追踪”方法被用来鉴定蛋白质结构域中哪些对其功能有很重要作用的结合位点。基于这个目的，在所有已知的蛋白质家族成员中，那些不发生变化的或几乎不变化的氨基酸可以定位在一个家族成员的三维结构模型上。这个工作完成以后，发现最保守的氨基酸常常成串出现在蛋白质表面，如图 3-40（A）中对 SH2 结构域的描述（讨论见专题 3-2）。这些成串的结构一般是配体结合位点。

SH2 结构域是一个模块，其功能是介导蛋白质-蛋白质相互作用。含有此结构域的蛋白

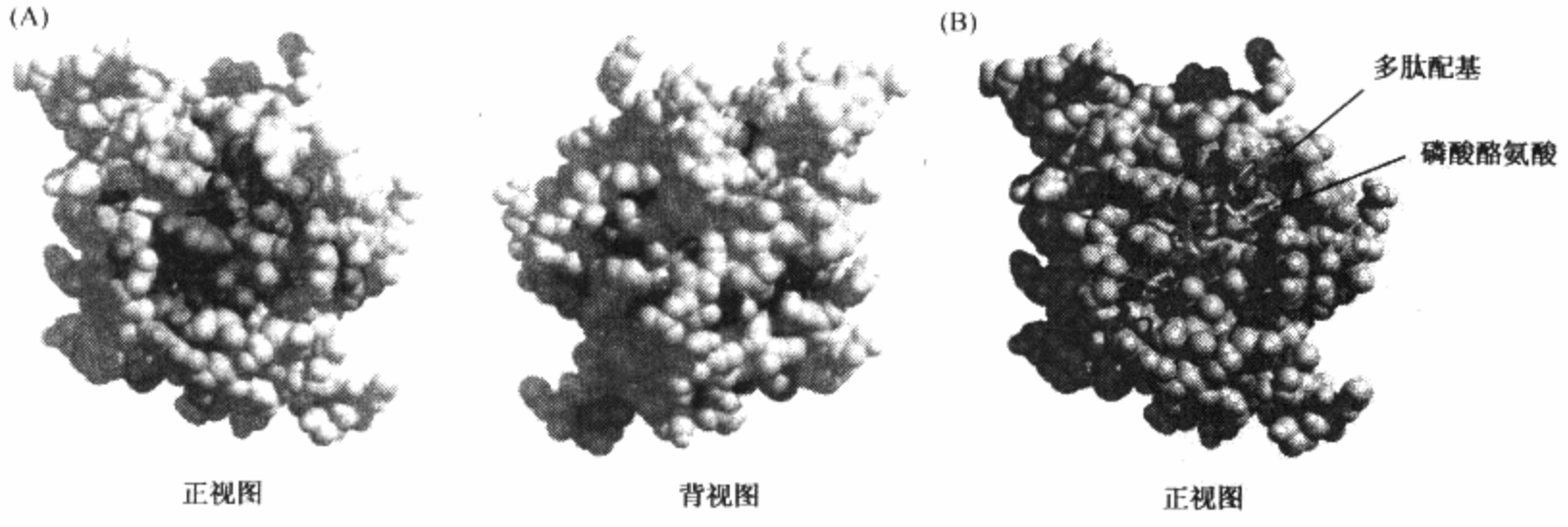


图 3-40　运用进化轨迹来研究 SH2 结构域

（A）SH2 结构域空间填充模型的正视图和背视图。蛋白质表面进化保守的氨基酸用黄色表示，面向蛋白质内部的氨基酸用红色表示。（B）SH2 结构域结合多肽链的结构。这里和配体距离在 0.4nm 的氨基酸用蓝色表示。配体中两个关键的氨基酸用黄色表示，其他的用紫色表示

质可以和另一个含有磷酸化酪氨酸的氨基酸序列的蛋白质结合，见图 3-40（B）。SH2 结构域家族中识别和磷酸化位点的氨基酸在进化过程中是改变最慢的。因为突变是一个随机的过程，这个结果因为偏好淘汰。所有有机体在进化过程中如果 SH2 结构域发生变化，使 SH2 结合位点失活，则会破坏其功能。

随着基因组测序的深入，发现了很多新的未知功能的蛋白质家族。在一个蛋白质家族成员的重要结合位点在三维结构上定位以后，上述的进化追踪可以用于了解这类蛋白质的功能。

蛋白质通过多种类型的接触面相互结合

蛋白质相互结合至少通过三种方式。很多情况下，一个蛋白质表面的一部分可以和第二个蛋白质多肽链（“绳”）上伸长的环相结合［图 3-41（A）］。这种表面-绳的相互作用，可以使 SH2 结构域识别第二个蛋白质磷酸化位点的环状结构，它也可以使一个蛋白质激酶识别它将磷酸化的另一个蛋白质（见下文）。

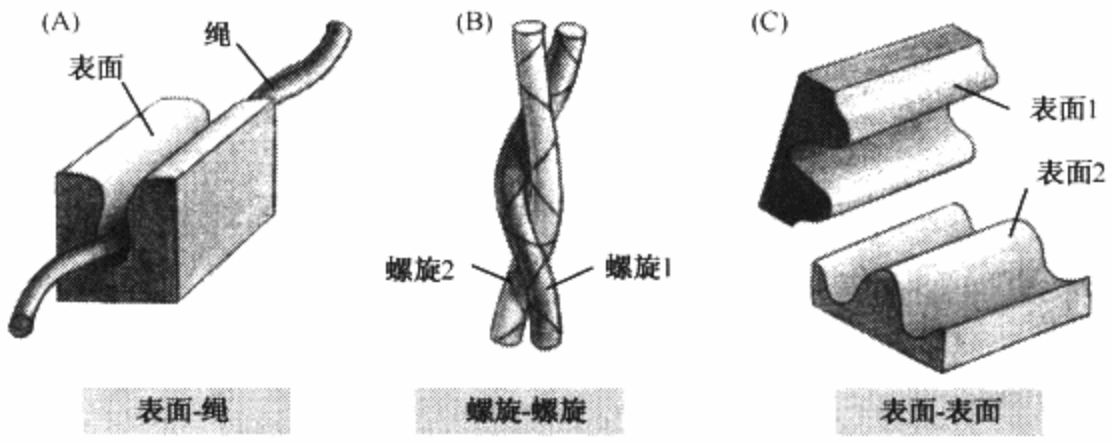


图 3-41　蛋白质相互结合的三种方式

只显示了两个蛋白质的相互作用的部分。（A）一个蛋白质的刚性表面可以和第二个蛋白质伸展的环形肽链（绳）结合。（B）两个 α 螺旋结合产生 coiled-coil 结构。（C）两个互补的刚性表面可以连接两个蛋白质

第二种类型的蛋白质-蛋白质相互作用是两个蛋白质上各有一个 α 螺旋，它们配对形成 coiled-coil［图 3-41（B）］。在第 7 章中，很多基因调控蛋白质家族的蛋白质采用这种作用方式。

最常见的蛋白质相互作用是两个刚性结构的精确配对［图 3-41（C）］。这种相互作用是非常紧密的，因为完全匹配的两个表面之间会形成大量的弱的相互作用。同样是这个原因，这种表面的相互作用可能是非常特异的，可以使一个蛋白质在胞内的成千上万种蛋白质中选择唯一的目标蛋白质。

抗体的结合位点是高度可变的

所有的蛋白质都必须和其相应的配体结合才能执行多种功能。这种高度选择性结合的能力在抗体家族中更加突出，这将在第 24 章中讨论。

抗体，或者免疫球蛋白，是免疫系统在应答外源物质（如入侵的微生物的表面分子）的过程中产生的蛋白质。每一种抗体可以和其相应的靶分子紧密结合，这样可以直接使靶标失活或者降解此分子。一个抗体分子和其靶标（**抗原**）的识别是高度特异的。我们可能会遇到上亿种不同的抗原分子，这就要求我们能够产生相应的这么多的抗体。

抗体是 Y 形的分子，有两个相同的结合位点和相应的抗原分子表面的位点互补。对抗体结合抗原分子的位点的仔细研究表明，它们是多肽链上一些环伸出到两个并列的蛋白质结构域的末端而形成（图 3-42）。不同抗体的抗原结合位点的多样性的产生是由于这些环的长度或者氨基酸的序列的改变造成的，抗体分子的基本结构并没有发生变化。

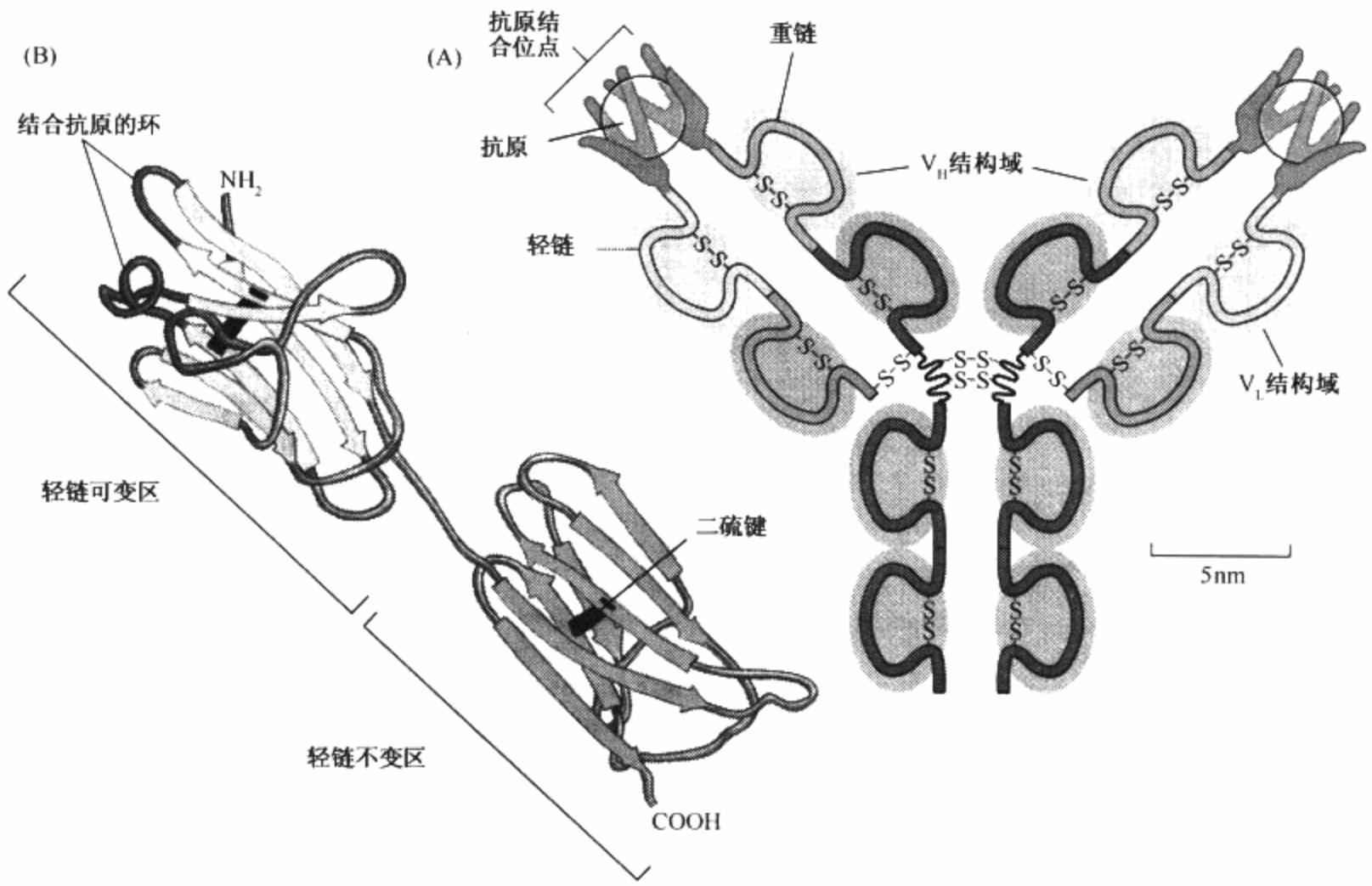


图 3-42　抗体分子

（A）典型的抗体分子是 Y 形的，有两个相同的抗原结合位点，“Y”的每个臂上一个。蛋白质是 4 条多肽链（两条相同的重链和两条相同的轻链）由二硫键连接起来。每条链由很多不同的免疫球蛋白结构域组成，以蓝色或灰色阴影表示。当一个重链的可变区（V_H）和一个轻链的可变区（V_l）结合时就形成了抗原结合位点。在不同的抗体分子中该结构域在序列和结构上都是变异最大的。（B）轻链的丝带模型显示了 V_L 结构域和抗体（红色）是最接近的。它们形成类手指环形的一半的结构，折叠在每个抗原分子周围（A）

这种环状结构更易捕获其他分子。它通过大量的化学基团围绕在一个配体周围，使得蛋白质可以以非常弱的键与其相作用。因此，蛋白质通常采用环的结构作为配体结合位点。

结合能力由平衡常数来衡量

分子在细胞中由于它们随机的热运动，相互碰到的频率是非常高的。当两个相互碰撞的分子的表面不匹配时，它们之间不会形成非共价键，两个分子会以其相互碰撞的速度快速地分开。在另一种极端的情况下，当形成了很多非共价键时，其解离会持续很长的时间（图 3-43），在细胞中当生物学功能需要分子长时间保持结合状态的话，就会形成强的相互作用，如 RNA 分子和蛋白质分子一起形成如核糖体这样的亚细胞结构。

任何两个分子相互结合的作用力可以直接测量。例如，设想一群相同的抗体分子突然遇

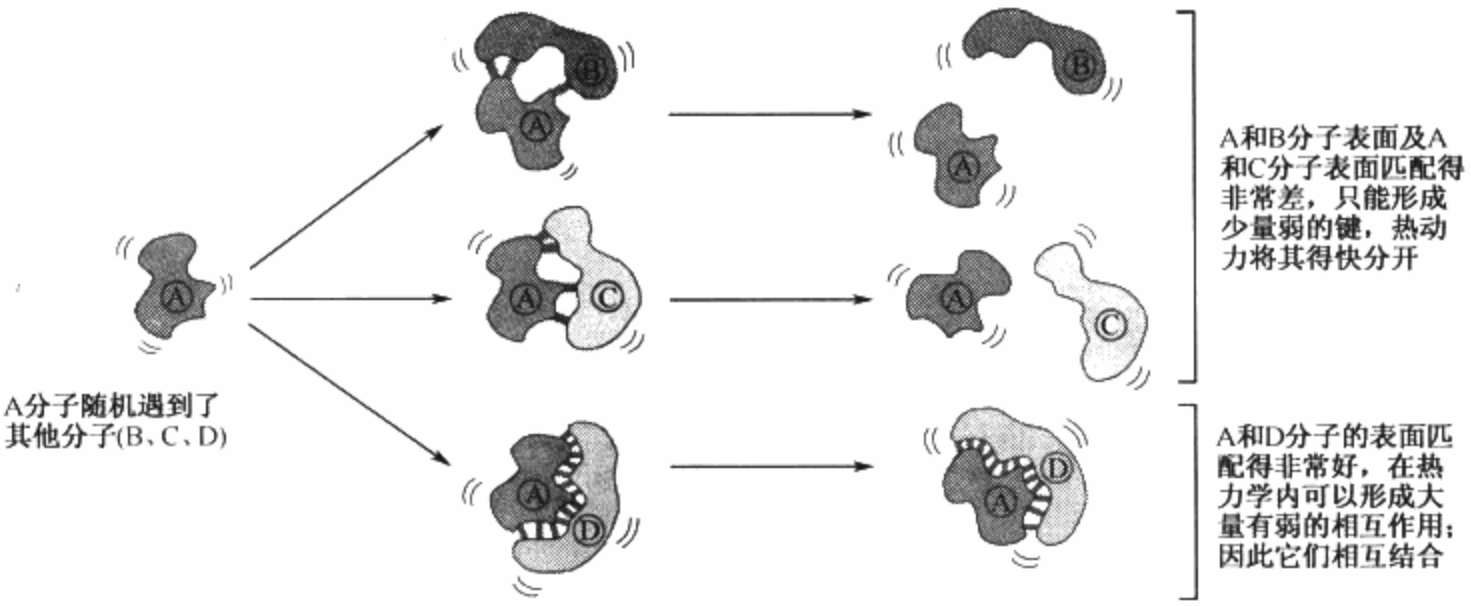


图 3-43　非共价键如何介导大分子的相互作用

到分散在体液中的一群配体分子。配体分子会不时地结合到抗体上形成抗体-配体复合物。那么抗体-配体复合物会增加，但不是无限制的增加：随着时间的延长，热运动每个复合物会被破坏，这是第二个过程，而且会变为主导。最后，所有的抗体、配体分子会达到一个稳定的水平，或者平衡，每秒结合的数量和去结合（解离）的数量是相等的。

根据配体、抗体和抗体-配体复合物平衡时的浓度，我们可以方便地计算**平衡常数**（K）——结合的强度［图 3-44（A)］。结合力越强，平衡常数越大，这是一个直接测量结合态和自由态自由能差异的方法［图 3-44（B)］。图 3-44（C）的例子表明改变很少的非共价键就有可能对蛋白质相互作用产生巨大的影响。

我们以抗体结合配体为例来说明平衡态下的结合能力，这个原理可以适用于所有的分子及其配体的相互作用。接下来将要讨论酶，它首先和其配体结合，接着催化分子之间共价键的断裂或者形成。

酶是高效性和高度专一性的催化剂

很多蛋白质只需要和其配体相互结合就可以发挥其生物学功能。例如，一个肌动蛋白分子只要和其他肌动蛋白结合就可以形成肌动蛋白纤维。但是，有些蛋白质，与配体结合只是其发挥功能的第一步。**酶**就是这样一类蛋白质，它是一类大而且非常重要的蛋白质家族。正如在第 2 章所说，酶是很重要的分子，它可以决定细胞内所有形成或者断裂共价键的化学反应。它们结合到一个或多个配体，或称为**底物**，快速地将其变为一个或多个化学修饰的产物。酶可以加快反应速度，通常以上百万的效率而其自身不发生任何变化，即它们作为催化剂在可以控制的条件下催化胞内共价键的形成或者断裂。酶的这种有组织的催化胞内化学反应的特点使得细胞可以维持下去，使生命得以存在。

根据酶促反应的反应性质可以将酶分类（表 3-1）。每一类中的每种酶都是非常特异的，只催化一种类型的反应。例如，己糖激酶只负责将磷酸基团加到 D-葡萄糖上，而对于其光学异构体 L-葡萄糖没有任何作用；凝血酶只催化一种类型的血液蛋白质特异的精氨酸和其相邻的甘氨酸的断裂等。在第 2 章已经详细讲述过，酶是以团体的方式起作用，一个酶的产物成为下一个酶的底物。这样会形成代谢途径中一个非常精密的网络，为细胞提供能量以及产生细胞需要的大分子和小分子（图 2-35）。

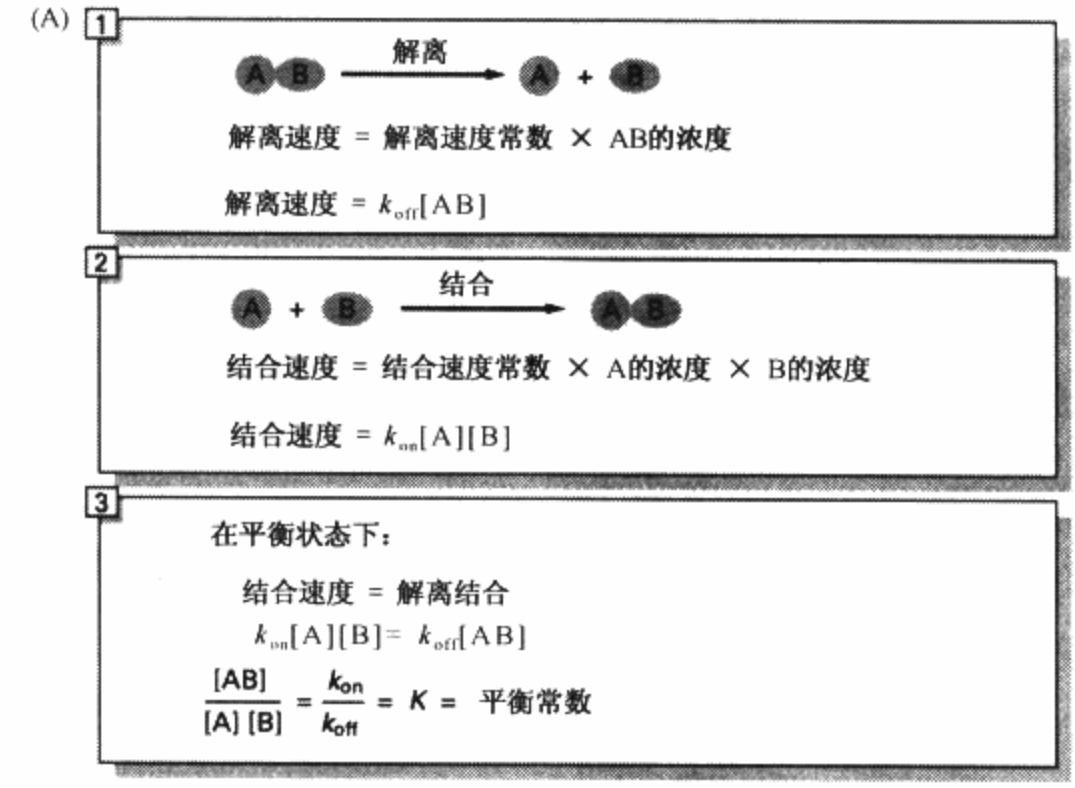


图 3-44　结合能和平衡常数之间的关系

（A）A 和 B 分子以及 AB 复合物之间的平衡由专题 3-1 和专题 3-2 中的两个相反的反应来维持。如果 A＋B＝AB 的平衡常数是 10^{10}，则在平衡态，如果平衡常数小于 10^8，这意味着会失去 2.8kcal/mol 结合能，或 2～3 个氢键，这会有 A 和 B 如果要发生反应，分子之间必须发生碰撞，结合速率和其自身的浓度［A］×［B］是呈正比的（方括号表示浓度）。如专题 3-3 所示，结合和解离的速度的比例等于反应的平衡常数（K）。（B）专题 3-3 的平衡常数是反应 A＋B⟷AB，值越大，表明 AB 之间的结合越强。每 1.4kcal/mol 自由能的释放，平衡常数提高 10。（C）存在或缺少弱的化学键在生物反应中的巨大影响的一个例子。

平衡常数的单位是 L/mol；对于简单的反应，它也叫做亲和常数或结合常数，K 以 K_a 表示。K_a 的倒数叫做解离常数，K_d（单位 mol/L）。

表 3-1　一些常见的酶的类型

酶	催化的反应
水解酶	通常用来表示催化水解反应的酶
核酸酶	通过水解核苷酸之间的键来破坏核酸
蛋白酶	通过水解氨基酸之间的键来降解蛋白质
合成酶	通常用来表示在合成代谢中连接两个小分子形成一个新的分子的酶
异构酶	催化一个分子内键的重排
聚合酶	催化聚合反应如 DNA 和 RNA 的合成
激酶	催化将磷酸基团加到分子上。蛋白激酶是一类重要的激酶，将磷酸基团加到蛋白质上
磷酸酯酶	催化从分子上水解去除磷酸基团
氧化还原酶	通常用来表示催化氧化还原反应的酶，一个分子发生氧化，而另一个分子发生还原。这种类型的酶也叫氧化酶、还原酶和脱氢酶
ATP 酶	水解 ATP。很多具有多功能的蛋白质有 ATP 酶的活性（作为其部分功能），如动力蛋白、肌球蛋白和膜转运蛋白钠钾泵

注：酶的名称通常以“-ase”结尾，但除了一些如胃蛋白酶（pepsin）、胰蛋白酶（trypsin）和凝血酶（thrombin）和溶菌酶（lysozyme），其命名在 19 世纪末期就被大家所接受。酶的名称通常表明其底物和催化反应的本质。如柠檬酸合成酶催化乙酰 CoA 和草酰乙酸合成柠檬酸。

底物的结合是酶促反应的第一步

对于一个催化化学反应的蛋白质（酶）来说，结合其相应的底物是一个很重要的前提。最简单的情况，我们以 E 代表酶，S 代表底物，P 代表产物，基本的反应是 E+S→ES→EP→E+P。根据这个反应式，在一段给定的时间内，一个酶分子可以加工的底物的量是有限的。如果底物的浓度增加，产物形成的速率也会增加到一个最大的值（图 3-45）。在这一点酶被底物所饱和，反应的速率（V_{max}）只取决于酶可以以怎样的速度加工底物。这个由酶浓度区分的最大速率叫做转换数量。转换数量通常是每秒每个酶分子催化 1000 个底物分子，现在已知转换数量可以为 1～10 000。

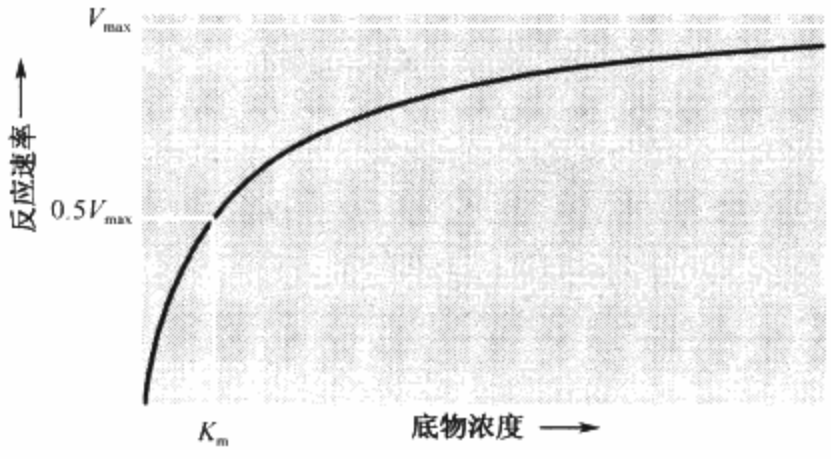


图 3-45　酶动力学

酶促反应的速度（V）随着底物浓度的增加而不断升高，直到达到一个最大值（V_{max}）。在这个点上，酶分子上的所有底物结合位点已经被占满，反应的速度由酶表面的催化过程所限制。对于大多数酶来说，反应速率达到最大值的一半（K_m）时的底物浓度用来测量酶和底物的亲和性，K_m 越大结合能力越弱

另外一个用来描述酶促反应的动力参数是 K_m，它是当酶促反应速度达到最大反应速度一半（$0.5V_m$）时的底物浓度（图 3-45）。K_m 越低意味着酶促反应达到最大反应速度时的底物浓度越低，它还可以反应酶对底物的亲和力，高的 K_m 值对应低的亲和力。这个描述酶促反应特征的方法在专题 3-3 中有详细说明。

专题 3-3　用来研究酶的一些方法

为什么要分析酶的动力学?

酶是已知的特异性和效率最高的催化剂。了解其机制可以为新药发现、大规模工业化生产化学物质、了解细胞和器官的化学特性提供一个重要的工具。详细研究由纯化的酶催化的化学反应的速率，更详细的是当外界条件变化，如底物、产物、抑制剂以及调节配体的浓度变化时，化学反应速率的变化，这使得生物化学家能精确地计算每个酶是如何工作的。例如，通过图 2-73 中糖酵解中的 ATP 合成的反应，我们可以知道这种重要的酶促反应途径。

在这个专题中，我们介绍了酶动力学的重要领域，从中我们可以得到很多细胞内的化学反应特征。

稳定状态酶动力学

很多酶只有一个底物，它们与之结合，并按照图 3-50（A）的方式对其进行加工产生产物。在这种情况下，反应可以这样表示

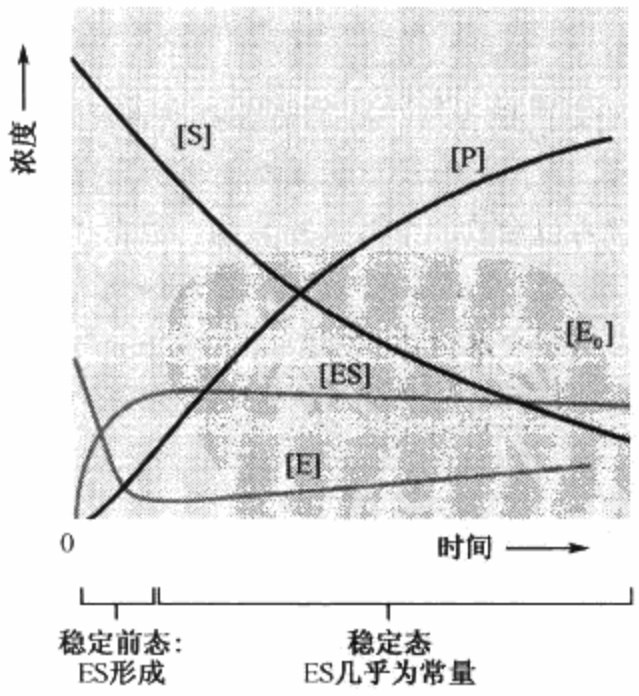
$$E+S \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} ES \xrightarrow{k_{cat}} E+P$$

这里我们假设逆反应，即 E+P 产生 EP 和 ES 的反应发生的概率很低，可以被忽略。在这种情况下，我们知道反应的速率 V 可以表示为

$$V=k_{cat}\ [ES]$$

[ES] 表示酶底复合物的浓度，k_{cat} 是转换数，即一个速度常数，表示的是每秒每分子酶加工的底物分子的数量。

但是 [ES] 的值如何和我们已知的浓度相关，即酶的总浓度 [E_0] 和底物的总浓度 [S]。当酶和底物混合时，[ES] 的浓度会快速的从零升到一个稳定状态，如下图所示



在这个状态，[ES] 是一个常数，所以

ES 分解 k_{-1} [ES] $+k_{cat}$ [ES] 的速度

=ES 合成 k_1 [E][S] 的速度

或者，因为自由的酶分子的浓度 [E] 等于 [E_0] − [ES]

$$[ES]=\left(\frac{k_1}{k_{-1}+k_{cat}}\right)[E][S]$$

$$=\left(\frac{k_1}{k_{-1}+k_{cat}}\right)\left([E_0]-[ES]\right)[S]$$

重排得到 K_m

$$\frac{k_{-1}+k_{cat}}{k_1}$$

得到

$$[ES]=\frac{[E_0][S]}{K_m+[S]}$$

或者 $V=k_{cat}$ [ES]，我们得到著名的 Michaelis-Menton 方程（米氏方程）

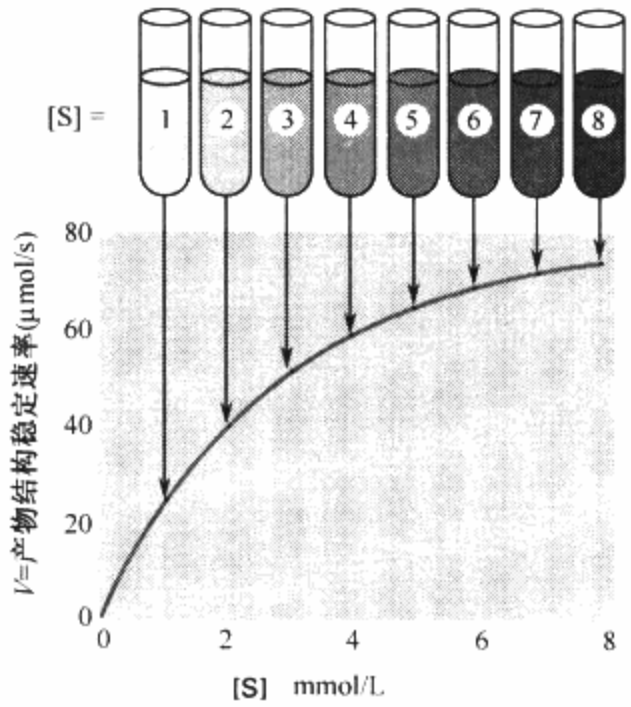
$$V=\frac{k_{cat}[E_0][S]}{K_m+[S]}$$

随着 [S] 的不断增加，所有的酶分子会和底物结合到稳定状态；在这一点上，当 $V=V_{max}=k_{cat}$ [E_0] 时，反应会达到最大速率 V_{max}。

$$V=\frac{V_{max}[S]}{K_m+[S]}$$

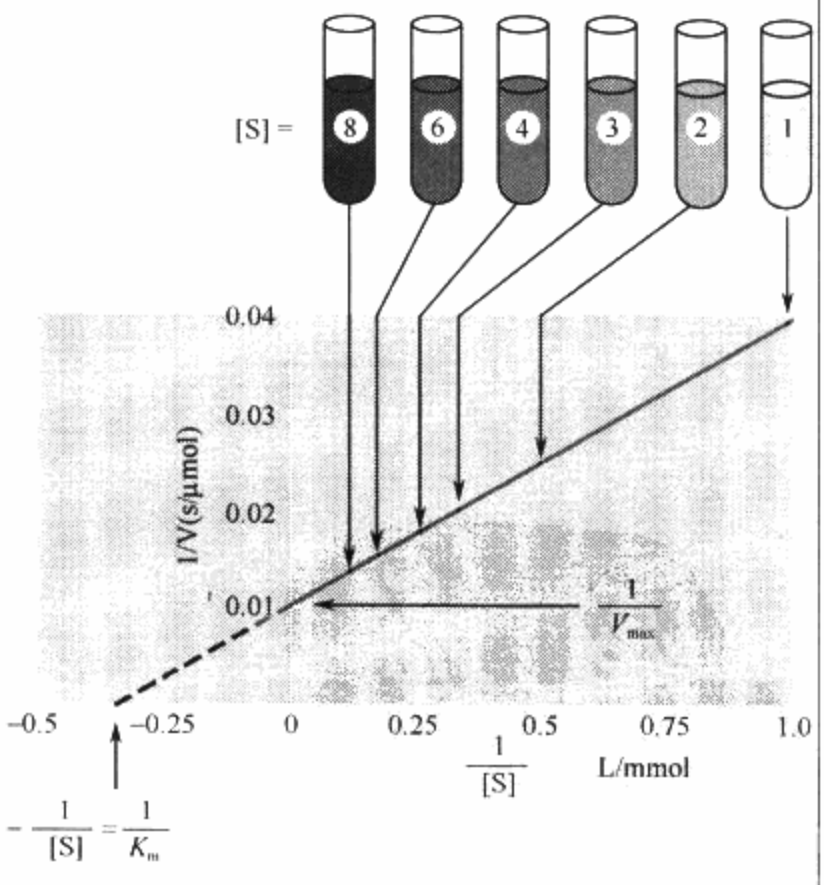
双倒数曲线

典型的 V 对［S］根据米氏动力学作图可得到如下图形。从这个图中我们不能直接清楚地得到 V_{max} 和 K_m 的值。



为了从这些数据中得到 V_{max} 和 K_m，可采用一个双倒数曲线，在这个图中米氏图被重新排列，所以 $1/V$ 可以对 $1/$［S］作图。

$$1/V=\left(\frac{K_m}{V_{max}}\right)\left(\frac{1}{[S]}\right)+1/V_{max}$$



K_m，K_{cat} 和 K_{cat}/K_m 的意义

如前所述，K_m 可以用来表示底物和酶的亲和力：当 $V=0.5V_{max}$ 时，$K_m=$［S］。一般来说，K_m 值小意味着和底物的结合能力强。

k_{cat} 表示酶的转换数。在低的底物浓度下，当［S］$\ll K_m$ 时，大部分酶处于自由状态。我们可以认为［E］=［E_0］，所以米氏方程变为 $V=k_{cat}/K_m$［E］［S］。因此，k_{cat}/K_m 的比例等于自由酶和自由底物反应的速率。

比较同一种酶的不同底物的 k_{cat}/K_m 值，或者是两种酶与其相应的不同底物的值，被广泛地用来作用测量酶反应效率的常数。

简单地说，在这个专题中，我们讨论了只有一个底物时酶的动力学，如溶菌酶。大多数酶有两个底物，其中一个通常是活性载体分子，如 NADH 或者 ATP。

相似的，但是更为复杂的是分析这种酶的动力学——底物的结合是有顺序的，而且在反应通路中存在共价中间物（图 2-73）。

某些酶是扩散限制的

某些酶的 K_m、k_{cat}、k_{cat}/K_m 如下所示：

酶	底物	k_{cat}/s	K_m /(mol/L)	k_{cat}/K_m /[s/(mol/L)]
乙酰胆碱酯酶	乙酰胆碱	1.4×10^4	9×10^{-5}	1.6×10^8
过氧化氢酶	H_2O_2	4×10^7	1	4×10^7
延胡索酸酶	延胡索酸	8×10^2	5×10^{-6}	1.6×10^8

因为酶和底物发生反应的前提是两者必须有碰撞，所以 k_{cat}/K_m 的最大值受碰撞速率的限制。如果每次碰撞可以形成一个酶底复合物，根据扩散理论我们可以算出 k_{cat}/K_m 的值为 $10^8\sim10^9$ s/（mol/L），这样所有接下来的步骤可以迅速发生。因此，如乙酰胆碱酯酶和延胡索酸酶被称作“完美的酶”，它们在每一次与底物的碰撞中都会将底物转换为产物。

酶通过选择性的稳定转换状态加速反应

酶催化的化学反应可以达到极高的反应速率，比其他任何合成的催化剂的效率要高很多。这么高的效率归功于很多因素。第一，酶可以增加催化位点局部的底物浓度，并且可以使所有参与化学反应的合适的原子处于正确的方向。更重要的是，催化作用直接决定了某些结合能量。底物分子在形成最终反应产物的过程中通过一系列中间态改变其构象和电子分布。达到最不稳定的**过渡态**所需要的自由能称为活化能，它是决定反应速率的主要因素，酶对于过渡态的底物有较高的亲和力。因为这种亲和力减少了到达过渡态所需要的能量，通过降低活化能，酶可以快速地催化化学反应（图 3-46）。

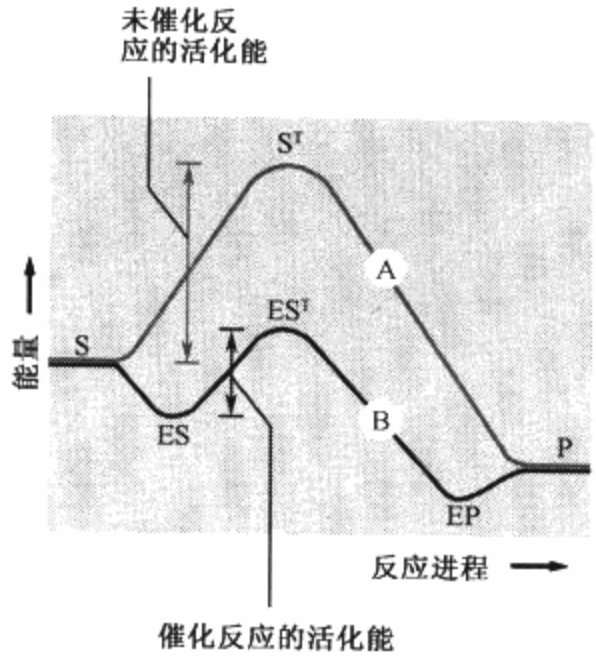


图 3-46　酶通过降低活化能加速反应
通常未催化的反应（A）和酶催化的反应（B）需要经过一些过渡态。过渡态的高能量（S^T和ES^T）决定了活化能并限制了反应速率。（S=底物；P=反应产物）

最重要的关于稳定过渡态可以增加反应速率的证据来自于抗体的产生是一个类似于酶的过程。试想一下，水解一个氨基，它和蛋白质中连接两个相邻氨基酸的肽键类似。在水溶液中，根据图 3-47（A）的机制水解一个氨基是非常慢的。在中间态，或过渡态，羰基碳原子与 4 个原子相连而处于一个四面体的角落上。在产生单克隆抗体过程中，抗体会和一个不稳定的四面体中间产物的稳定的类似物结合，这样就实现了抗体分子作为酶的功能［图 3-47（B)］。因为这种催化性抗体和四面体中间物结合并使之稳定，所以它使自发水解氨基的速度增加了 10 000 倍。

酶可以同时产生酸催化和碱催化作用

图 3-48 比较了 5 种酶的自发反应速率和相应的酶催化的反应速率。催化速度可达到 $10^9 \sim 10^{23}$ 。酶是比催化性的抗体要好的催化剂。酶不仅和过渡态紧密结合，而且它们可以精确定位那些因为直接参与形成或断裂共价键而改变其电子分布的原子。将肽键置于一个没有酶，而只有强酸或者强碱的环境中，肽键可以被水解，详细说明见图 3-49（D)。酶是非常独特的，它可以同时利用酸催化和碱催化作用，这需要酸性残基和碱性残基不发生相互结合（在溶液中它们通常都会结合)，紧密地结合到蛋白质自身的刚性骨架上［图 3-49 (D)］。

酶和其相应的底物需要精确的配对。利用遗传工程，对酶活性中心的一个微小的改变都可能产生严重的影响。例如，用天冬氨酸代替一个谷氨酸后，使起催化作用羧基仅仅偏离了 1Å（相当于一个氢原子的半径)，但是却使这个酶的活性降低了 1000 倍。

溶菌酶揭示了酶是怎样发挥作用的

为了解释酶是如何催化化学反应的，我们以溶菌酶为例，它在鸡蛋清、唾液、眼泪和其他分泌物中作为一个天然的抗菌物质。**溶菌酶**催化细菌细胞壁的多糖水解。由于细菌细胞壁处于一定的渗透压下，所以少量多糖的水解就会导致细胞壁的裂解和细胞的崩溃。溶菌酶是

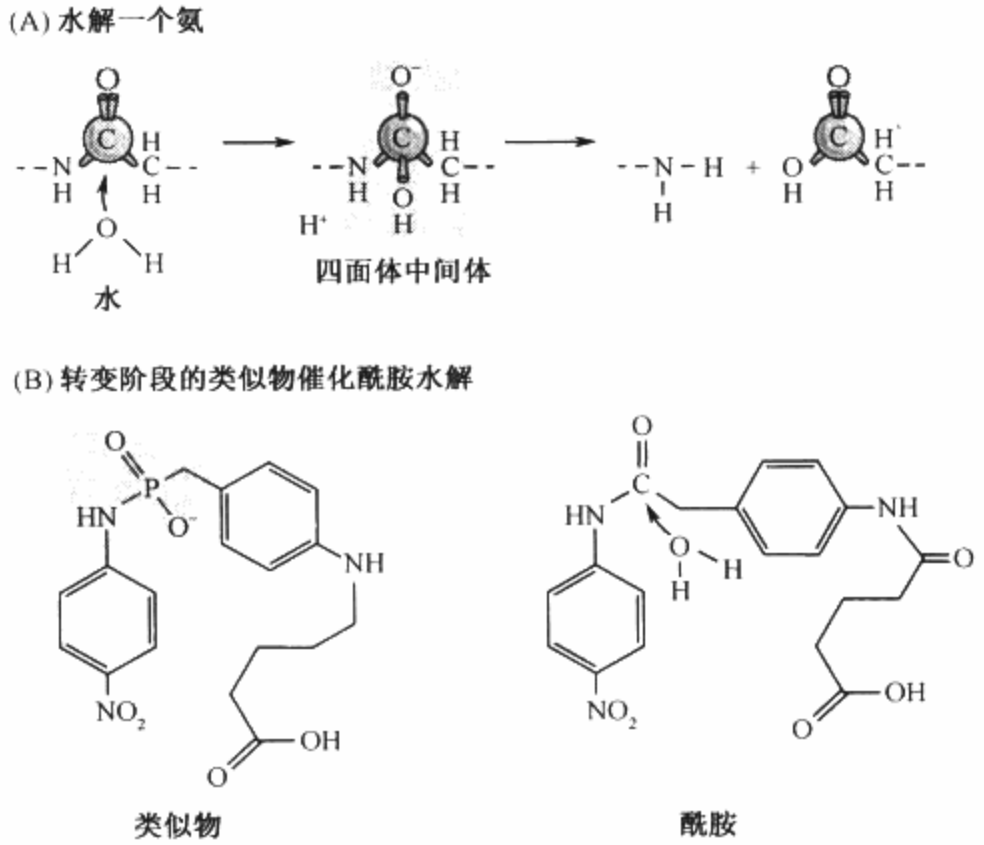


图 3-47　催化性抗体

由抗体稳定的过渡态可以产生酶分子。（A）水解氨基的反应经过了 4 步中间反应到达高能量的过渡态。（B）左边的分子共价交联到一个蛋白质上，可以作为抗原产生抗体和图中分子的黄色区域紧密结合。因为该抗体分子也可以和（A）的过渡态结合，所以它具有酶的功能，可以有效地催化右边分子氨基的水解

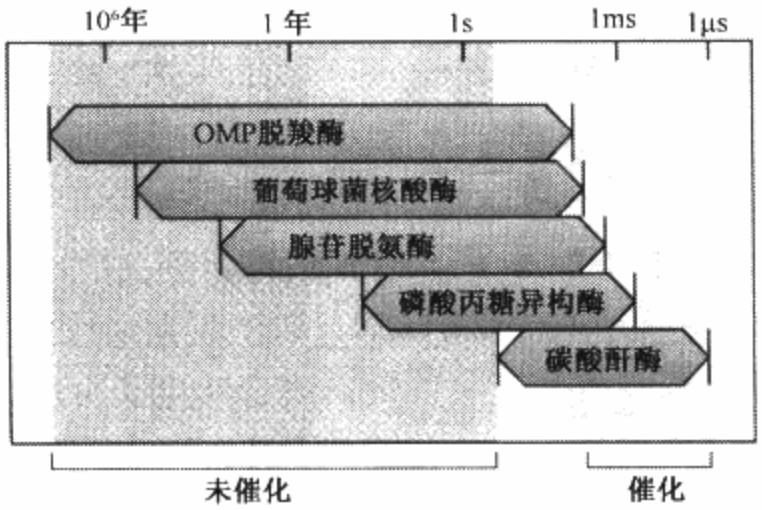


图 3-48　由 5 种不同的酶催化的反应

（修改自 A. Radzicka and R. Wolfenden，*Science* 267：90～93，1995.）

一个分子质量小，且非常稳定的蛋白质，很容易被大量分离。因为这些原因，人们对它的研究很透彻，它是第一个用 X 射线衍射阐明其原子结构的酶。

溶菌酶催化的反应是水解反应：一分子水加入到多糖链上的两个相邻糖基的一个键上，而使键断开（图 2-19）。这个反应是能量偏好的，因为切断的多糖链的自由能比完整链的自由能要低。但是，纯的多糖链在水中可以稳定存在很长时间而不发生降解。这是因为存在一个能量屏障，已在第 2 章讲述过（图 2-46）。一个水分子要想破坏连接两个糖分子的键，必须要求多糖链弯曲成一个特殊的形状——过渡态，键周围的原子必须改变其几何构象和电子分布。由于这个扭曲，在反应发生前必须通过随机碰撞产生大量的活化能。在室温下水溶液中，碰

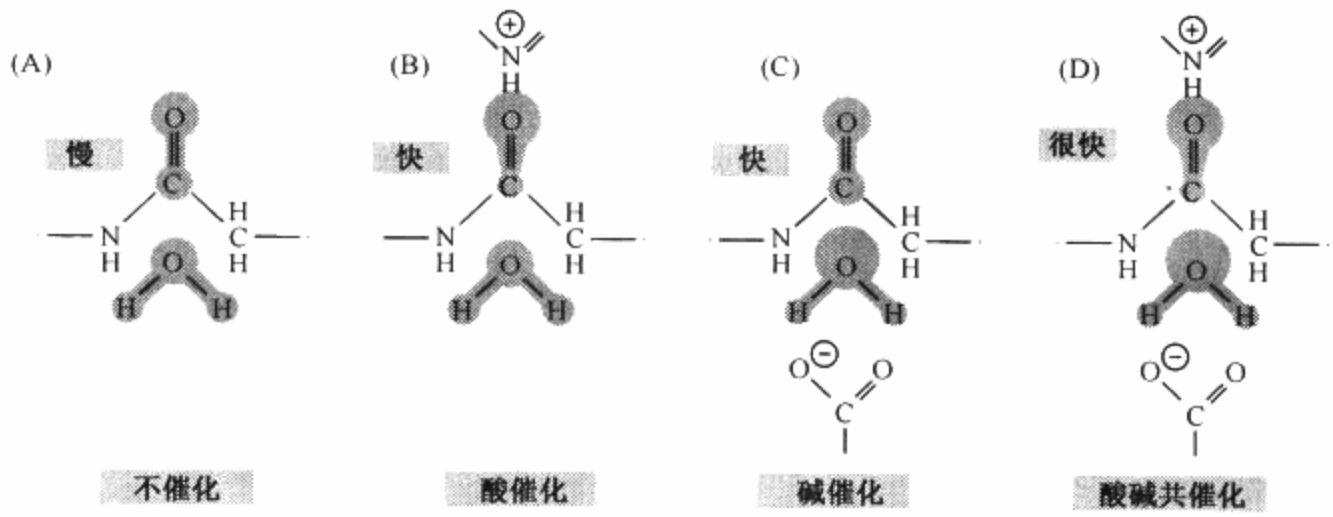


图 3-49　酸催化和碱催化

（A）图 3-47（A）中表示初始的未催化的反应，蓝色表示水和羰基上的电子分布。（B）酸可以为其他原子提供一个质子（H^+）。通过和羰基氧配对，酸促使电子从羰基碳上离开，使该原子易于和被攻击的水分子上的带负电的氧原子相互吸引。（C）碱易于夺取 H^+。通过和攻击的水分子上的氢原子配对，碱促使电子朝向水分子中的氧，使其成为羰基碳易攻击的基团。（D）当所有的原子在表面处于一个正确的位置时，酶分子可以同时起到酸催化和碱催化的作用

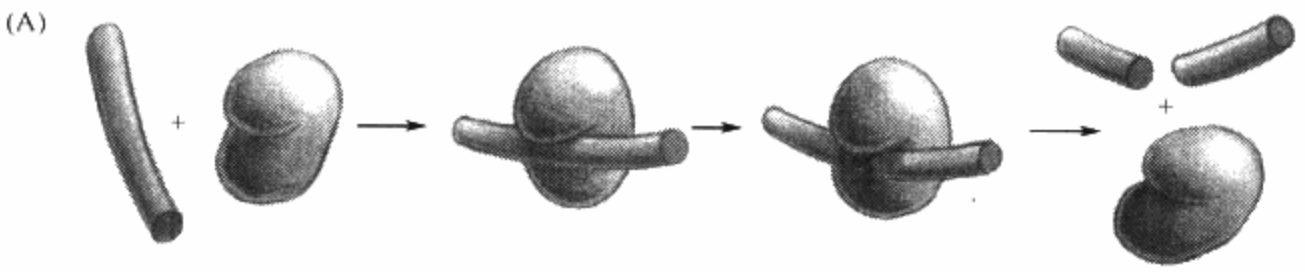


图 3-50　溶菌酶催化的反应

（A）溶菌酶（以 E 表示）催化其底物（S）多糖链的切割。酶首先和链结合形成一个酶底复合物（ES），接着催化多糖链骨架特异共价键的切割，形成酶-产物复合物（EP）并迅速解离。切断的链（产物 P）的释放使酶可以作用于其余的底物分子。（B）溶菌酶分子结合短的切割前的多糖链的空间填充模型［（B）经 Richard J. Feldmann 许可］

撞的能量不会超过活化能。结果，水解发生得极其缓慢或者根本不发生。

当多糖链结合到溶菌酶上后这个状况就会发生极大的改变。由于其底物是一个聚体，溶菌酶的活性中心形成一个长形凹穴，可以同时容纳 6 个糖分子。随着多糖结合到溶菌酶上形成一个酶底复合物，酶通过插入 1 分子水到糖-糖键之间而切断多糖。产物链的释放非常快，使得有更多的酶进行循环反应（图 3-50）。

溶菌酶结合其相应底物的化学本质和抗体结合抗原的化学基础是一致的——形成多个非共价键。但是，溶菌酶和其多糖底物的结合采用一种特殊的方式，将断裂的键涉及的 2 个糖分子中的一个发生扭曲，偏离其原始的、正常的、稳定的构象。将断裂的键同时也靠近酶的两个酸性氨基酸（一个谷氨酸和一个天冬氨酸）活性中心。

溶菌酶活性中心的微环境就形成了这样一个合适的条件可以减少水解所需要的活化能。图 3-51 显示了 1967 年提出的酶促反应的过程。

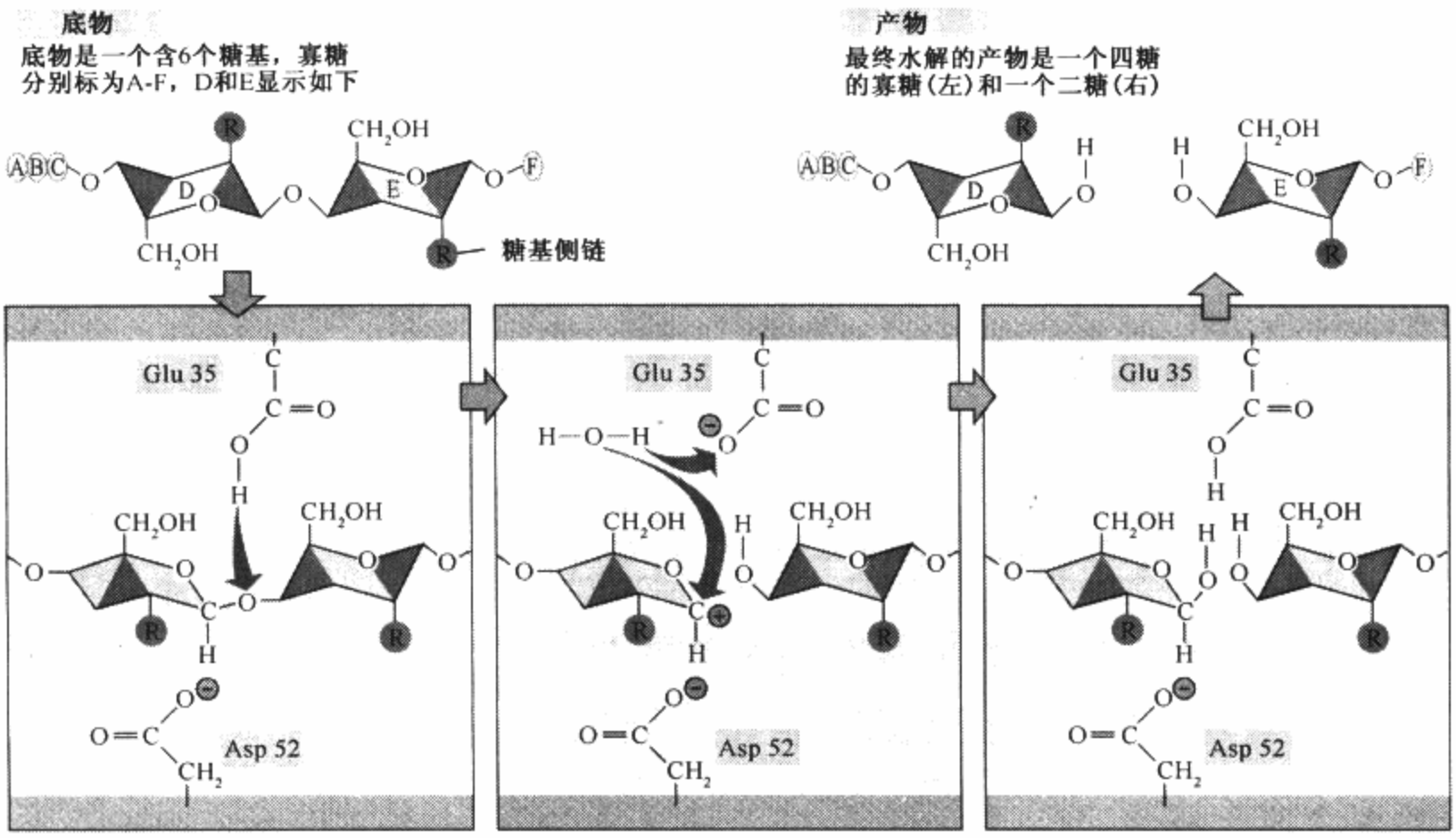


图 3-51　经典的在溶菌酶活性中心发生的事件

上面左边和右边的图分别描述了自由的底物和自由的产物，而其余三个图描述的是在酶活性中心发生的一系列事件。注意在酶-底物复合物中糖 D 的构象变化；如中间专题所示，我们假设此糖在不稳定的过渡态中带正电。但是，最近的研究发现这个标准教科书的观点是不正确的。相反，Asp52 与碳原子形成的共价键是带正电的。这个共价键在下一步随着水分子参与到此反应中，立即被破坏

1. 酶感受底物以后使某些重要的参与化学反应的化学键弯曲，在反应中途，致使底物的构象成为一个能量较高的过渡态。

2. 虽然酶周围的环境是处于一个中性的 pH 条件，酶活性中心的谷氨酸的酸性侧链通过提供一个高浓度的质子化的 H^+（酸催化）加速水解反应。

3. 天冬氨酸的负电荷可以稳定正电荷的过渡态（碱催化）。

如图注中所示，溶菌酶反应的经典观点需要修改。特别是，在过渡态会形成一个四面体中间产物，该中间产物是由天冬氨酸暂时共价地加到切割位点处的碳原子上形成的。

相似的原理可以被应用到其他酶促反应中，通过降低活化能来加速其所催化的反应。涉及两个或两个以上反应物时，活性中心会作为一个模板或者模型，使参与反应的底物聚集在一个合适的方向［图 3-52（A）］。正如溶菌酶，酶的活性中心包含精确定位的原子，通过带电基团改变底物的电子分布而加速化学反应［图 3-52（B）］。底物与酶的结合也会改变底物的形状，弯曲的化学键使底物向过渡态转变［图 3-52（C）］。最后，如溶菌酶，很多酶在催化反应时会在底物和酶的一个侧链之间形成共价键。接下来的反应步骤会使侧链逐渐恢复成其原始的状态，所以在反应结束后，酶没有发生变化（图 2-73）。

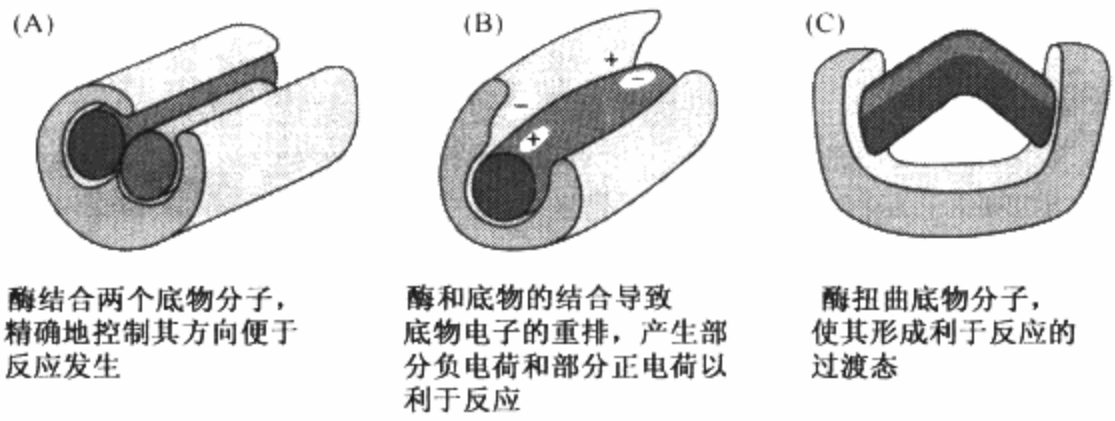


图 3-52　酶催化的一些策略

（A）和底物形成一个精确的排列；（B）反应中间物的电荷稳定性；（C）改变底物的键角增加反应的速率

与小分子的高亲和性赋予了蛋白质额外的功能

虽然我们已经强调过蛋白质的多功能性由氨基酸链来执行，但是在很多情况下仅依靠氨基酸自身是不够的。正如人使用工具来增强其手的能力，蛋白质也会利用非蛋白质小分子来执行其功能，而这些仅利用氨基酸是很难或根本不可能完成的。例如，信号受体蛋白视紫红质（rhodopsin）是视网膜中光受体细胞产生的，可以通过嵌入该蛋白质的小分子视黄醛（retinal）来感应光信号［图 3-53（A）］。当视黄醛吸收一个光子以后，其构象会发生变化，这个改变使得蛋白质诱发一系列级联酶促反应，最终导致电信号被大脑接收。

另外一个含有非蛋白质成分的蛋白质是血红蛋白（图 3-23）。1 分子的血红蛋白包含 4 分子环状的血红素，每一分子有一个中央的铁原子［图 3-53（B）］。血红素使血红蛋白（血液）成为红色。通过铁原子它可以和氧气可逆结合，使得血红蛋白在肺部吸收氧气，在组织中释放氧气。

有时候小分子与蛋白质共价结合，并且这种结合是永久性的，因此它变成蛋白质分子自身的一个部分。在第 10 章中我们会看到蛋白质通过和脂分子共价结合而锚定在细胞膜上。而位于细胞表面的膜蛋白或分泌到细胞外的膜蛋白通常和糖或多糖共价结合。

酶中的小分子或金属原子通常和其活性中心密切相关，并且可以帮助其催化功能。例如，羧肽酶（carboxypeptidase）可以切断多肽链，其活性中心有一个紧密结合的锌离子。在酶切割多肽链的过程中，锌离子和底物的一个原子暂时形成一个化学键，帮助水解反应。在其他酶中，小的有机分子有类似的作用。这些有机分子通常叫做**辅酶**。一个典型的例子是生物素，它帮助羧基（—COO—）从一个分子向另一个分子转移（图 2-63）。生物素通过和羧基形成暂时的共价键参与化学反应，它比任何构成蛋白质的氨基酸更适合这个反应。人自身不能合

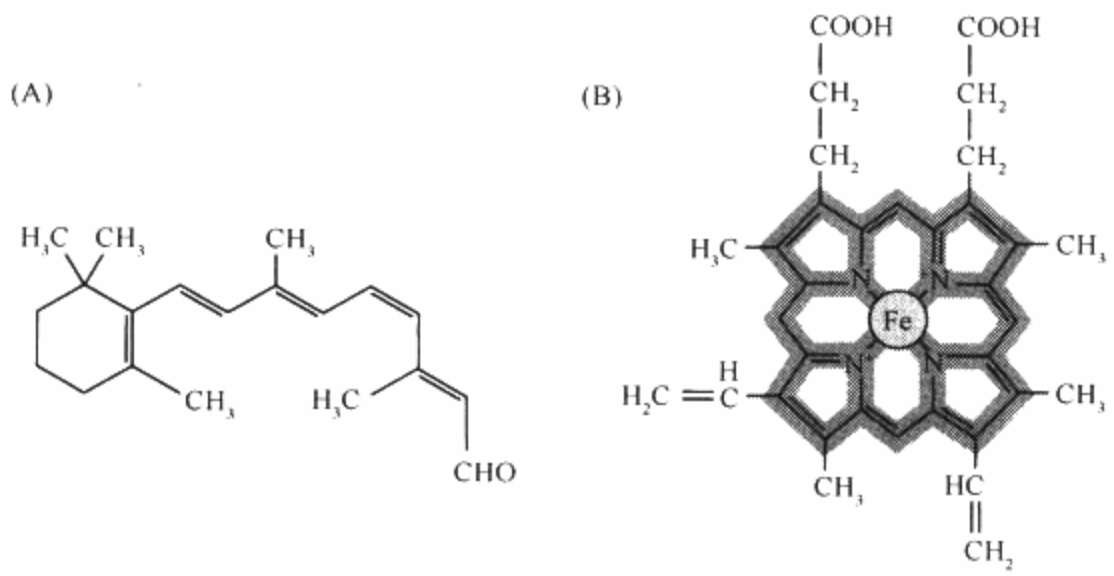


图 3-53　视黄醛和血红素

（A）视黄醛的结构。它是光敏感分子，在眼睛中与视紫红质结合。（B）血红素的结构。含有碳原子的血红素环是红色的，铁原子位于其中央是橙色的。一个血红素分子和血红蛋白的 4 条多肽链结合，图 3-23 中有该携带氧气的蛋白质的结构

成该物质，所以我们必须在饮食中补充，如很多其他的辅酶一样，生物素是一种维生素 A（表 3-2）。其余的维生素构成其他小分子，这些小分子是蛋白质的必要组分。例如，我们需要在饮食中补充维生素 A 来生成视黄醛（视紫红质的光敏感部分）。

表 3-2　很多维生素为人类细胞提供必需辅酶

维生素	辅酶	酶催化的反应需要这些辅酶
硫胺素(维生素 B_1)	焦磷酸硫胺素	乙醛的激活和运输
核黄素(维生素 B_2)	FADH	氧化-还原
烟酸	NADH,NADPH	氧化-还原
泛酸	辅酶 A	酰基的激活和运输
维生素 B_6	吡哆醛磷酸盐	参与氨基酸活化的反应
生物素	生物素	CO_2 的活化和运输
硫辛酸	硫辛酰胺	酰基活化;氧化-还原
叶酸	四氢叶酸酯	单碳基团的激活和运输
维生素 B_{12}	钴胺素辅酶	异构作用和甲基基团的运输

多酶复合物帮助增加细胞代谢速率

酶对化学反应的催化效率对维持生命是很关键的。实际上，细胞必须和不可避免的蛋白质相竞争，如果不注意，会使大分子越来越无序。如果所需的反应的速率没有侧链竞争反应的速率低，一个细胞很快就会死亡。通过测量 ATP 的消耗水平可以了解细胞代谢的速率。一个典型的哺乳细胞每 1 或 2 分钟即“转换”（如水解和磷酸化恢复）其 ATP 库。对于每个细胞，这种转换代表了每秒利用 10^7 个分子 ATP（或对于人来说，每分钟利用 1g ATP）。

因为酶催化反应的高效性，细胞中的代谢速率是非常快的。很多重要的酶的效率之高以致不需要另外的改进。限制反应速度的因素并不是酶催化反应的能力，而是酶与底物碰撞的频率。这种反应是扩散限制的。

如果酶促反应是扩散限制的，那么它的速度取决于酶和底物共同的浓度。如果一个反应的速度要求非常快，那么参与反应的中间代谢产物和酶的浓度都必须非常高。但是，试想一下，细胞中的反应如此之多，可以达到的底物的浓度是有限的。实际上，很多代谢物的浓度只有微摩尔（10^{-6} mol/L），而且大部分酶的浓度更低。那么怎么样维持细胞内快速的代谢呢？

答案在于细胞组分的空间结构组织。通过把参与一系列代谢反应的酶聚在一起形成一个大的蛋白质集合，如多酶复合物（图 3-54）就可以在不需要增加底物浓度的情况下提高反应速度。因为这种方式可以使酶 A 的产物直接运输到酶 B 上，依次类推，扩散的速度不需要限制，甚至当细胞中的底物浓度作为整体时仍然很低。这并不令人感到惊讶，这种多酶复合物是很常见的，它们参与到细胞代谢的各个方面，包括遗传中心 DNA、RNA 的加工，蛋白质的合成。实际上，真核细胞中很少种类的酶是自由扩散于溶液中的；大多数酶都有结合位点和其他功能相关的蛋白质在细胞的特定区域结合，提高其浓度和催化的反应的速度和效率。

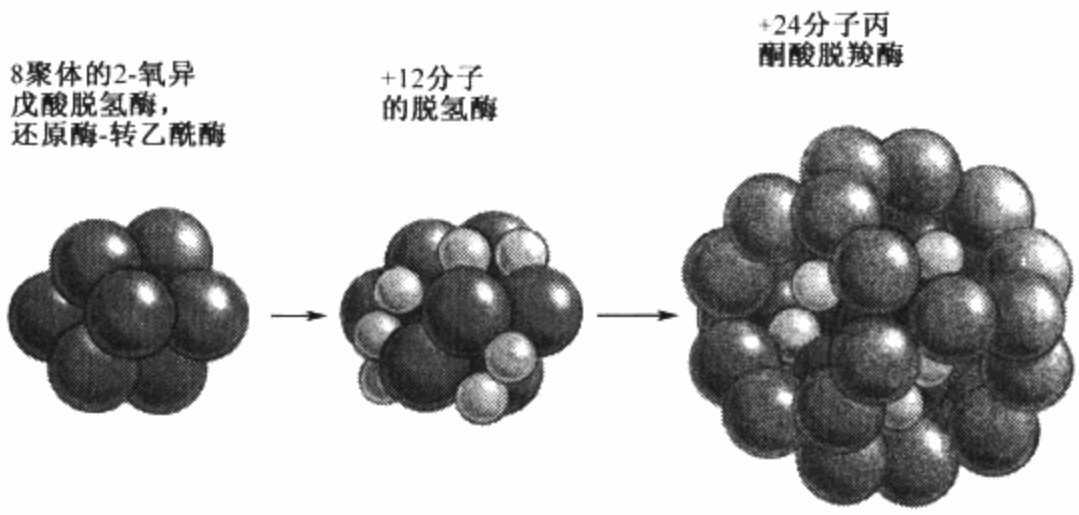


图 3-54　丙酮酸脱氢酶的结构

这个酶复合物催化丙酮酸转变为乙酰 CoA，是糖氧化为水和二氧化碳的中间反应。这个例子表明大的多酶复合物其反应中间物是如何直接从一个酶传递到另一个酶的

真核生物也有另一种方式来提高代谢速度，即利用细胞内的膜系统。这些膜可以将特定的底物及其相应的酶封闭在一个特定的细胞器内，如内质网或者细胞核。如果一个细胞器占有细胞内体积的 1/10，反应物在细胞器内的浓度比同一个细胞内不处于细胞器内的反应物浓度高 10 倍。因为反应与扩散的速度的关系，其速度会提高 10 倍。

酶的催化活性可以被调节

一个活细胞包含成千上万种酶，很多会在同一时间、细胞的同一空间起作用。通过其催化作用，这些酶产生了一个复杂的代谢通路网络，每种包含化学反应链，一个酶的产物是另一个酶的底物。在这个迷宫般的体系中，有很多支路不同的酶竞争同一个底物。这个系统是如此的复杂（图 2-88），所以需要一套精确的调控系统来调节每个反应在何时何地以何种速度发生。

调控在很多水平上起作用。一个水平是，通过调节编码这种酶的基因的表达水平来调控每种酶的分子数目（讨论见第 7 章）。细胞也通过限制每种酶在特定的由不同膜系统控制的亚细胞器中来调节酶促反应（讨论见第 12 章和第 14 章）。蛋白质降解的速度代表了另外一个重要的调控机制。但是最快的、最常见的调控方式是通过酶活性中心对特定分子的直接的、可

逆的应答。

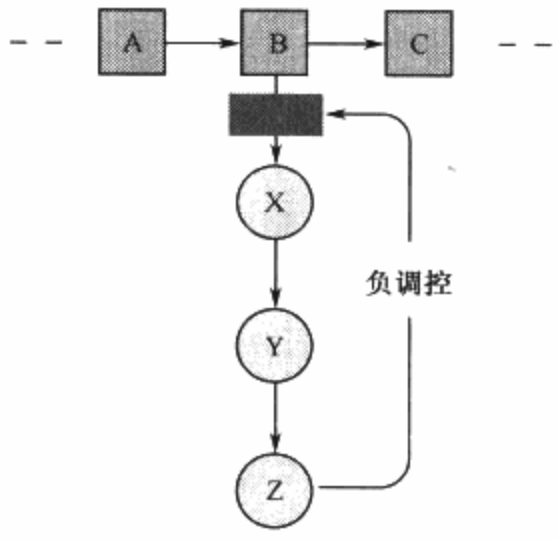


图 3-55　一条生物合成途径的反馈抑制
终产物 Z 抑制第一个酶，该酶是其合成中的特有的，因此控制了细胞内其产物自身的水平。这是一个负调控的例子

最常见的调控方式发生在当多个底物结合到一个酶的非活性中心的调控位点时，因此改变了酶催化底物转变的速度。在反馈抑制中，反应初期起作用的酶会被通路的后续产物所抑制。因此，当大量的最终产物开始积累时，这个产物会结合到反应的第一个酶上而降低催化反应速度，而限制了接下来的底物进入反应通路（图 3-55）。当通路交叉时，有很多由不同的反应终产物调节的控制点，每一种控制其自身的合成（图 3-56）。反馈抑制可以立即发挥作用，而当产物量降低时又很快地被逆转。

反馈抑制是一个负调控，它阻止了一个酶的激活。酶也有正调控，酶的活性可以被一个调节分子激活。正调控发生在当一个复杂的代谢通路上一条支路的产物激活了另一条通路上的酶的活性时。例如，ADP 的积累激活了很多参与糖氧化的酶的活性，加速了细胞将 ADP 转化为 ATP。

别构酶有两个或多个结合位点相互作用

反馈抑制最初让其发现者感到非常困惑的一个特点是：调节分子和酶底物的构象完全不同。这也是为什么这种调控称为*别构*（来自于希腊语 allos，意思是“其余的”或 stereos，意思是“固体的”或“三维”）的原因。随着我们对反馈抑制的深入了解，发现很多酶分子表面至少包含两个不同的结合位点，一个识别底物的**活性位点**，另一个识别调控分子的**调控位点**。这两个位点必须相互关联使发生在活性中心的催化反应被结合在蛋白质表面的另外一个位点的调控分子所影响。

一个蛋白质分子表面分隔的位点之间的相互作用依赖于蛋白质的构象变化：结合于一个位点会使蛋白质从一个构象变化为另外一个不同的构象。在反馈抑制中，一个抑制剂结合在一个蛋白质上会使蛋白质变为另外一个构象，其活性中心（位于蛋白质另外一个地方）失活。

大多数酶是别构酶。这种酶有两个或多个不同的构象，一个配体的结合可以使构象发生转变而改变其活性。这个效应不仅在酶中，在很多其他的蛋白质，如受体、结构蛋白和动力蛋白中也存在。别构调节不是一个神秘的事件：蛋白质的每种构象有不同的表面结构，当蛋白质的构象发生变化时，其配体结合特性也会发生变化。而且我们接下来将会讨论，每个配体结合在其紧密的构象上时，会稳定该结构，高浓度的配体可以使蛋白质的构象发生转变。

两个配体如果其结合位点是偶联的，将会相互影响各自的结合性

蛋白质上的配体结合性的影响来自一个基本的化学原理——**连锁**。假设一个蛋白质可以结合葡萄糖，也可以利用其表面的远程结合位点与另一个分子 X 结合。如果葡萄糖结合时会导致构象发生变化，而致使蛋白质表面的 X 结合位点也发生构象变化，而此构象是葡萄糖结合时构象变化的一部分，那么 X 和葡萄糖的结合位点就是偶联的。当两个配体结合别构蛋白的同一个构象时，根据基础热动力学理论，每个配体分子会增加蛋白质与另外一个配体的亲和性。因此，如果图 3-57 中蛋白质构象的变化会使之和葡萄糖更好地结合，那么它同样增加了 X 结合的亲和性，即 X 存在时，蛋白质和葡萄糖的结合比其不存在时会更加紧密。

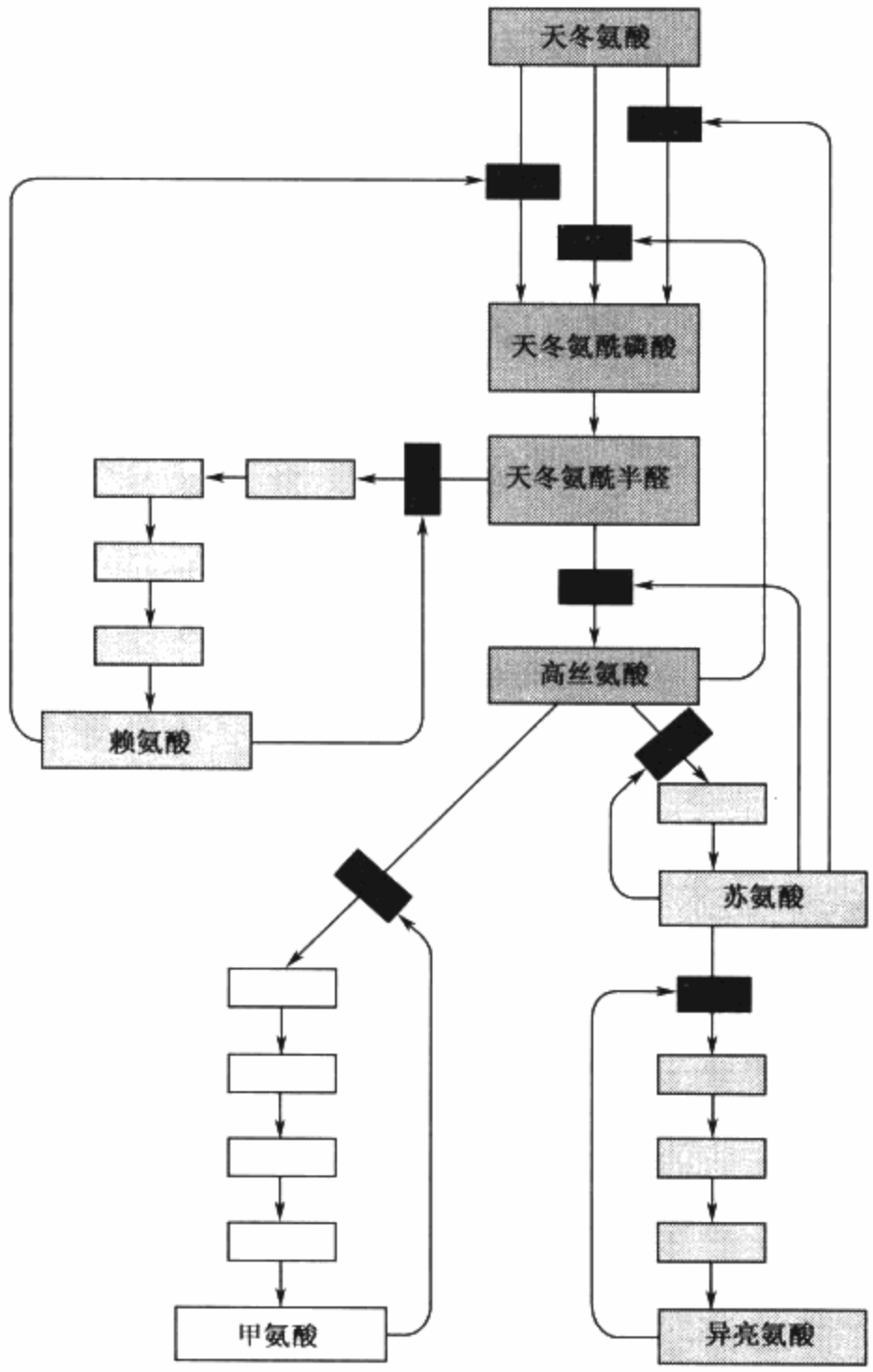


图 3-56　多重反馈抑制

在这个例子中出现的是细菌内 4 种不同的氨基酸生物合成的途径，红色的箭头表示哪个产物反馈抑制酶。每个氨基酸控制其自身合成特异的酶，因此可以控制其自身水平，以免浪费中间产物。产物也可以分别独立抑制对于所有合成反应的最初共同的反应；在这种情况下，三种不同的酶催化初始反应，每一种被不同的产物抑制

相反地，如果两个配体倾向于与蛋白质的不同构象结合，那么连锁理论就会起负调控。在这种情况下与第一个配体的结合，会破坏与第二个配体的结合。因此，如果葡萄糖结合时蛋白质构象的变化降低了蛋白质对 X 的亲和性，那么 X 的结合必然也降低了蛋白质对葡萄糖的亲和性（图 3-58）。连锁的关系跟量也有关，如果葡萄糖对 X 的结合有很大的影响，那么 X 对于葡萄糖的结合也有很大的影响。

图 3-57 和图 3-58 所示的关系存在于所有的细胞中。当我们回顾起来时，都觉得这是理所当然的。但是它是在 20 世纪 50 年代被发现的，在 60 年代的时候才有别构理论来详细描述，这在当时是一个革命性的发现。在这些例子中，X 分子都和催化活性位点相距较远，它需要葡萄糖或者其他结合在活性位点的配体不发生任何化学反应。正如我们刚才所知，酶是以这种方式被调节的，X 分子可以开启（正调控）或者关闭（负调控）酶促反应。根据这样的机

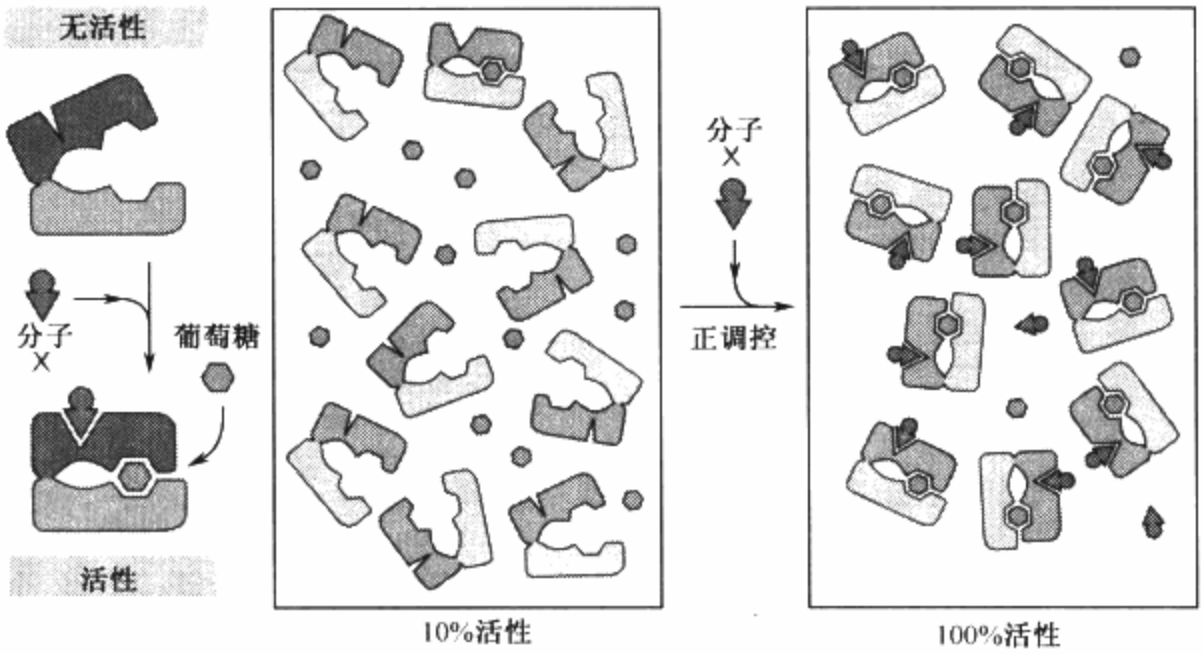


图 3-57　两个远的结合位点的构象偶联导致的正调控

图中，葡萄糖和 X 分子与一个有两个结构域的蛋白质的闭合构象很好地结合。因为葡萄糖和 X 分子使蛋白质趋向于其闭合的构象，所以每一种配体帮助其余的配体结合。因此，葡萄糖和 X 分子叫做协同结合

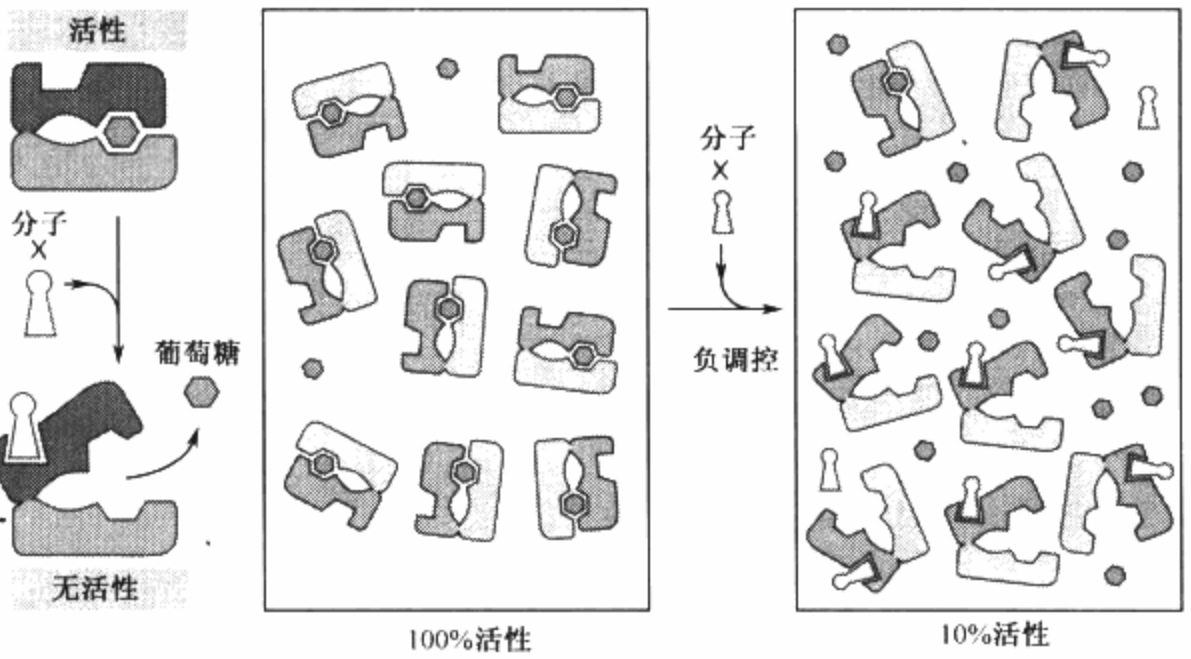


图 3-58　两个远的结合位点的构象偶联导致的负调控

该图和上图很像，但是在这里 X 分子易于和蛋白质的开放构象结合，而葡萄糖易于和蛋白质的闭合构象结合。因为葡萄糖和 X 分子使蛋白质趋向两个相反的构象（闭合和开放），一种配体的结合影响了另一种配体

制，别构蛋白作为一个常见的开关，允许细胞中的一个分子影响别的分子的命运。

对称的蛋白质分子的组装产生协同别构转换

一个酶的亚基如果被负反馈调控，随着抑制性配体的浓度增加 100 倍，其活性从 90%降低到 10%（图 3-59，红线）。这种类型的应答对于最佳的细胞调控并不是最迅速的，大多数酶通过配体结合到包含有对称亚基的分子上从而被调控。根据这种模型，一分子配体结合到一个位点上会导致亚基上所有的别构变化而影响其相邻的亚基，帮助它们结合到相同的配体上。结果，一个协同别构转换发生（图 3-59，蓝线），它利用细胞内配体浓度的微量变化来开启整个复合体从一个完全有活性的状态变为一个完全无活性的构象（或相反）。

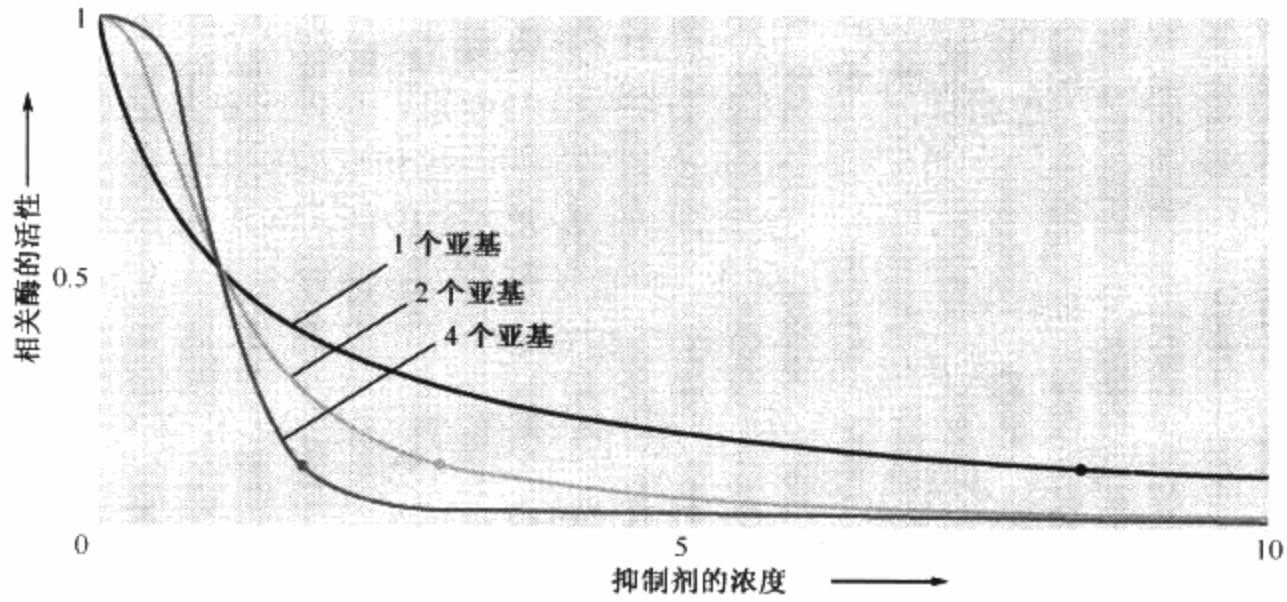


图 3-59　酶的活性与单亚基或多亚基别构酶抑制配体浓度的关系

单亚基酶（红线）其活性从 90%降低到 10%（曲线上由两个点表示）需要抑制剂的浓度增加 100 倍。根据一个简单的平衡关系 $K=[I][P]/[IP]$（P 是一个活性蛋白质，I 是抑制剂，IP 是结合了抑制剂的非活性蛋白质）计算酶活性。这个典型的曲线可以适用于任何简单的两个分子 AB 的相互作用。相反的是，多亚基别构酶以一种开关的方式应答配体的浓度变化：最直接的应答是由于如图 3-60 中的配体分子的协同结合。在这里，绿线表示两个抑制配体分子协同结合到有两个亚基的别构酶的理想结果，蓝线表示 4 亚基别构酶。从每条曲线上的两个点可以得出：越复杂的酶，其活性从 90%下降到 10%所需要的抑制剂的浓度变化范围小于单亚基酶

在协同的“全或无”转换中，最容易理解的是那些形成对称二聚体的酶。在图 3-60 中所示的例子中，第一分子抑制配体的结合是非常困难的，因为它的结合破坏了二聚体中两个单体之间的相互作用。但是第二分子抑制配体的结合则会较容易一些，因为其结合保留了对称二聚体中单体-单体之间的相互作用（几乎使酶的活性完全丧失）。

配体与更大的复合物结合能产生更灵敏的，如我们将要讨论的由 12 条多肽链构成的酶。

原子水平上了解天冬氨酸转氨甲酰酶的别构转换

来源于大肠杆菌的天冬氨酸转氨甲酰酶是早期研究别构调节的一个酶。它催化嘧啶核苷酸 C、U 和 T 的最初合成：氨甲酰磷酸＋天冬氨酸⟶ *N*-氨甲酰天冬氨酸。这个反应体系的最终产物之一，三磷酸胞嘧啶（CTP），当其浓度达到一定水平时会结合到酶上关闭此酶促反应。

天冬氨酸转氨甲酰酶是一个由 6 个调节亚基和 6 个催化亚基组成的大的复合物。催化亚基以 3 个二聚体的形式存在，形成一个等边三角形；2 个三聚体互相面对面，中间由 3 个调节二聚体亚基作为桥将其连接。整个分子在两种构象［称为 T（紧密的）和 R 情况（松散的状态）之间进行协同的、全或无的别构转换，从而保持平衡。

底物（氨甲酰磷酸和天冬氨酸）结合到催化三聚体上，随着 CTP 分子的解离，促使天冬氨酸转氨甲酰酶转变为其有活性的 R 态。相反地，CTP 结合到调节亚基上随着底物的解离，促使酶变为无活性的 T 态。这 CTP 和底物之间的关系与图 3-58 中简单的别构蛋白的调控机制是相同的。但是由于竞争发生在有多个结合位点的对称分子上，酶采用了一个协同的别构调节来迅速调节酶的状态变化：底物积累时（形成 R 态）开启酶活，CTP 积累时（形成 T 态）关闭酶促反应。

运用生物化学手段和 X 射线晶体衍射阐明了很多别构转换的细节。每个调节亚基有两个结构域，CTP 的结合促使两个结构域的构象接近，它们的功能就像一个杠杆旋转着两个催化三聚体，推动它们一起向 T 态称为（图 3-61）。当此发生时，相对的催化亚基上形成氢键，

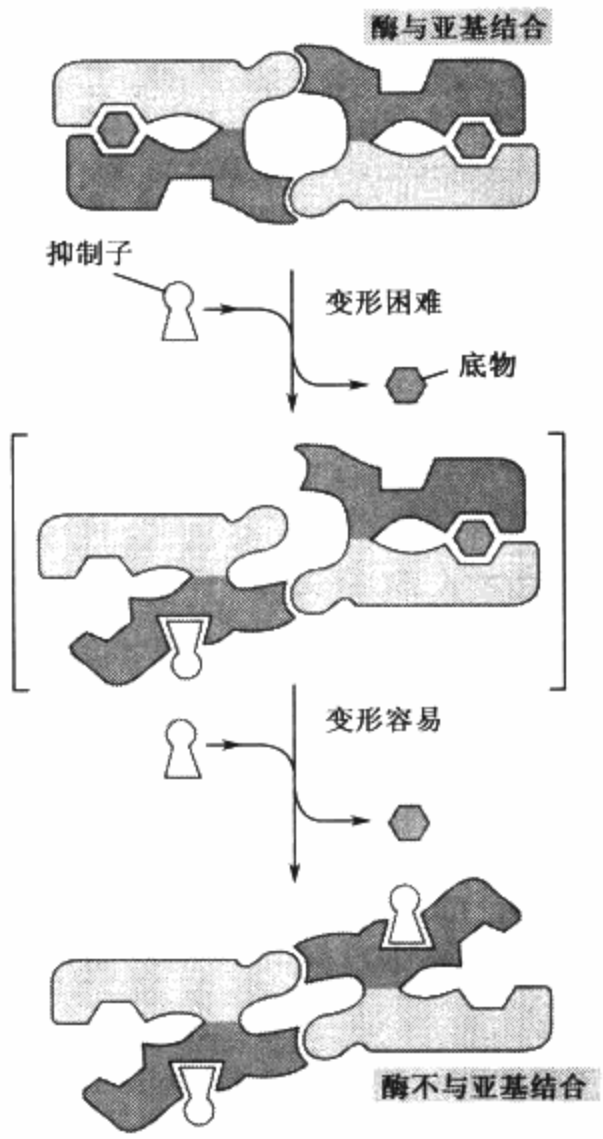


图 3-60　由两个相同亚基组成的酶的协同别构转换

该图揭示了一个亚基的构象怎样影响其邻居的构象。一分子抑制配体（黄色）和酶的一个亚基的结合是比较困难的，因为它改变了该亚基的构象而破坏了酶的对称性。但是，一旦构象发生了变化，重建两个亚基之间的对称相互作用需要的能量促使第二个亚基很容易和抑制配体结合并且也发生同样的构象变化。因为第一个配体分子的结合增加了其余亚基和同样配体的结合，酶对配体浓度变化的应答比对单亚基酶的应答要灵敏得多（图 3-59）

帮助每个催化亚基扩大裂缝来形成一个催化活性中心，从而破坏了底物的结合位点（图 3-62）。增加底物的浓度就会有相反的效果，与催化亚基裂缝的结合促使形成 R 态，导致与上述相反的构象变化。介于 R 态和 T 态之间的构象是不稳定的，酶在 R 态和 T 态之间来回转换，形成这两种构象的混合物，其比例取决于 CTP 和底物的相对浓度。

磷酸化会导致蛋白质的很多变化

酶不仅仅通过与小分子的结合被调控。常被真核细胞用来调控蛋白质功能的第二种方法是通过共价连接一个磷酸基团到其氨基酸链上。这种磷酸化可以在两个方面影响蛋白质的活性。

首先，因为磷酸基团带有两个负电荷，酶促反应给蛋白质加上一个磷酸基团后会造成蛋白质构象的变化，如吸引带正电的氨基酸成团。这样会影响蛋白质表面配体的结合，通过别构效应可能会极大地改变蛋白质的活性。通过第二个酶去掉磷酸基团会使蛋白质变回其原来的构象而恢复其初始活性。

其次，磷酸基团的加入会形成可以被其他蛋白质结合位点直接识别的结构的一部分。如前所述，某些称为模块的小的蛋白质结构域，在大的蛋白质中出现的频率很高。大多数这些模块为这些蛋白质与其他蛋白质磷酸化的多肽链结合提供了结合位点。SH2 结构域就是这样的一个模块，它结合到一个短的包含酪氨酸磷酸化位点的肽链上［图 3-40（B）］。很多其他类型的模块以特殊的方式识别磷酸化的丝氨酸或苏氨酸侧链。因此，蛋白质的磷酸化和去磷酸化对调节蛋白质复合物的装配与去装配有很大的作用。

通过这些作用，可逆的蛋白质磷酸化调控了真核细胞多种蛋白质的有活性的结构以及细胞定位。实际上，这种调控是很广泛的，一个典型哺乳动物细胞内的 10 000 个蛋白质中有 1/3 应该在合适的时间被磷酸化，很多不只包含一个磷酸基团。可以想像，蛋白质磷酸基团的添加和去除通常发生在对信号的应答过程中，而这又决定了细胞的命运。例如，真核细胞分裂过程中一系列复杂的事件按此运作（讨论见第 17 章），通过级联的蛋白质磷酸化事件，而使很多介导细胞-细胞相互作用的信号从细胞膜释放到细胞核。

真核细胞中有一大类蛋白激酶和蛋白磷酸（酯）酶

蛋白质的磷酸化是指酶催化一分子 ATP 末端的磷酸基团转移到蛋白的丝氨酸，苏氨酸或酪氨酸侧链的羟基上的过程（图 3-63）。这个反应由**蛋白激酶**催化，是单向的，因为 ATP 分子中的磷酸-磷酸键断裂产生 ADP 时会释放大量的自由能（讨论见第 2 章）。磷酸基的去除反

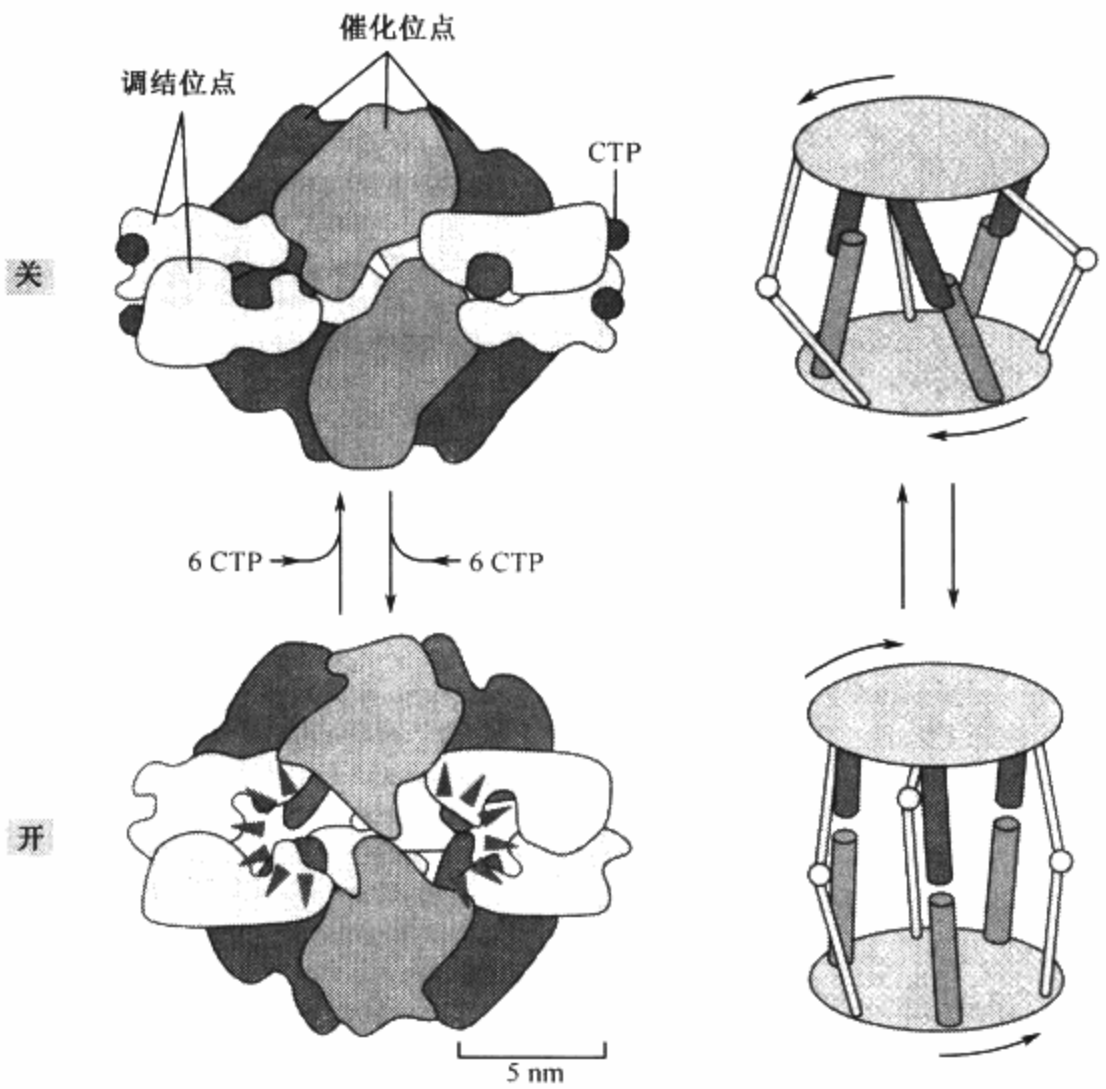


图 3-61　天冬氨酸转氨甲酰酶 R 态和 T 态的转换

该酶包括 6 个催化亚基和 6 个调节亚基，其无活性状态（T 态）和活性状态（R 态）结构已经用 X 射线衍射解出。当 CTP 浓度上升时，该酶因为反馈抑制而失活。每个调节亚基可以结合一分子的 CTP，它是该反应的终产物。通过负反馈调节，该反应会被抑制以防止在细胞中产生过多的 CTP（基于 K. L. Krause，K. W. Volz，and W. N. Lipscomb，*Proc. Natl. Acad. Sci.* USA 82：1643～1647，1985.）

应，或者叫做脱磷酸化，由**磷酸酯酶**催化。细胞内有大量不同的蛋白激酶，每一种磷酸化不同的蛋白质。同样也存在不同的蛋白磷酸酯酶；很多是高度特异的，只脱去一种或少数几个蛋白质的磷酸基团，而另一些可能作用的范围比较广泛，通过调节亚基结合到特异的底物上。蛋白质磷酸化状态的出现以及其活性取决于蛋白激酶和蛋白磷酸酯酶的相对活性。

真核细胞中使蛋白磷酸化的蛋白激酶属于一个大的酶家族，它包含一个由 250 个氨基酸组成的催化结构域（激酶）。家族中的成员在激酶序列的任何一端都有不同的氨基酸组成（图 3-12），通常有短的氨基酸序列插入到环区（图 3-64 中红箭头）。一些额外的氨基酸序列使每种激酶可以识别其磷酸化的蛋白质的特异位点，或者结合特殊的结构而使其可以定位到细胞内特殊的区域。蛋白质的其他部分允许每种酶的活性被精确的调控，如前所述可以根据不同的信号而做出相应的应答。

通过对比一个蛋白质家族中不同成员的氨基酸序列的差异，我们可以构建一个“进化树”，它可以反映出基因复制模式和导致蛋白质进化的变异方式。图 3-65 给出了蛋白激酶的进化树。并不奇怪，功能相关的激酶通常位于树分支中很近的位置。例如，在细胞信号转导中参与磷酸化酪氨酸的蛋白激酶聚集在进化树左边角落的顶端。其他的蛋白激酶可以磷酸化

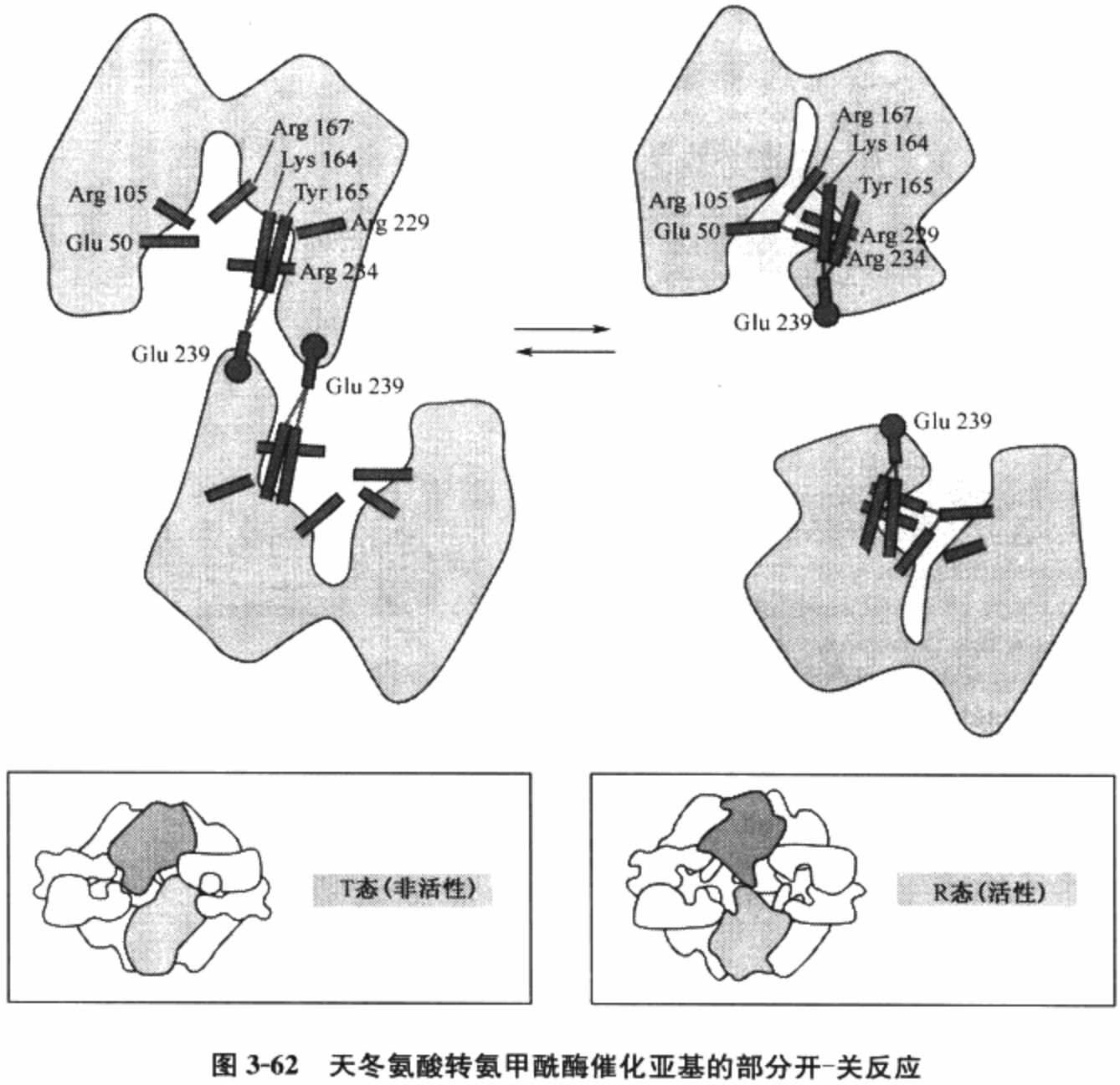


图 3-62　天冬氨酸转氨甲酰酶催化亚基的部分开-关反应

氢键作用力的变化在一定程度上决定了酶的构象在活性（黄色）和非活性状态的转换。氢键用细的红线表示。参与了亚基相互作用的氨基酸用红色表示，而形成酶活性中心的氨基酸用蓝色表示。最大的图表示酶内部的活性中心；方框的简图是从酶的外表面观察的酶的亚基（E. R. Kantrowitz and W. N. Lipscomb，*Trends Biochem. Sci.* 15：53～59，1990.）

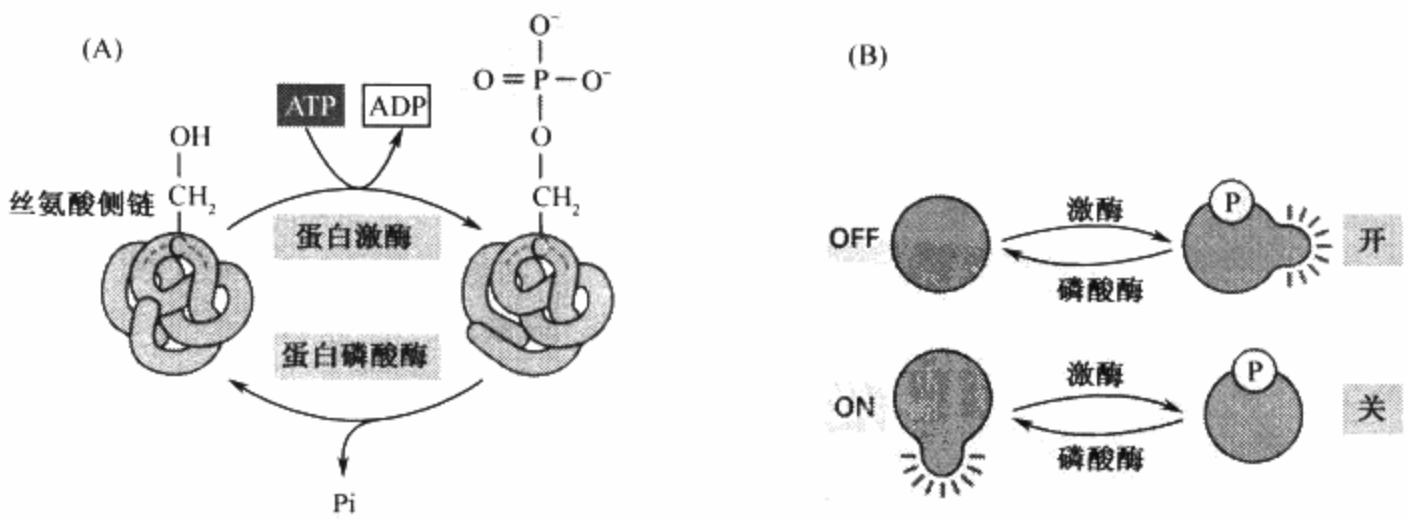


图 3-63　蛋白质的磷酸化

一个典型的真核细胞中的很多蛋白质都要进行磷酸化。（A）一般的反应是由蛋白激酶将 ATP 上的磷酸基团转移到靶蛋白的侧链上。蛋白磷酸酯酶则负责去除磷酸基团。在该例子中，磷酸被加到丝氨酸侧链上；在其他情况下，磷酸基团也可以加到苏氨酸或酪氨酸的羟基上。（B）由蛋白激酶介导的磷酸化可以增加或者降低蛋白质的活性，这主要依赖于磷酸化的位点和蛋白质的结构

丝氨酸或苏氨酸，很多成簇排列反应了其功能，如跨膜信号转导、胞内信号扩增、细胞周期调控等。

由于蛋白激酶和蛋白磷酸酯酶的活性，蛋白质上的磷酸基团会持续地颠转，快速加上或者去除。这样的磷酸化循环看起来似乎是很浪费的，但是它们对于磷酸化的蛋白质迅速地从一个状态转变为另一个状态是非常重要的：循环越快，一群磷酸化的蛋白质在应答外界会改变磷酸化速度的突然刺激时越快（图15-10）。磷酸化循环所需要的能量来自于ATP水解所释放的自由能，每次磷酸化消耗一分子的ATP。

Cdk和Src蛋白激酶的调控显示了一个蛋白质是如何作为一个微芯片起作用的

真核细胞内多种蛋白激酶形成一个复杂的信号通路网络，来协同调控细胞的活性、推动细胞周期、将外界的信号传递到细胞内。许多细胞外的信号需要被整合到细胞内，并进行扩增。一个蛋白激酶（和其他的信号分子）在整合的过程中作为输入-输出的引擎，或“微芯片”。在信号网络中，信号的输入主要由磷酸化控制，通过蛋白激酶和磷酸酯酶分别加上或去除磷酸基团来实现。

如果一个蛋白质可以在多个位点磷酸化，特异的磷酸基团可以激活蛋白，其他位点的可以让它失活。细胞周期依赖的蛋白激酶（Cdk）是这类信号途径引擎的一个代表。这类蛋白激酶磷酸化丝氨酸，它们是真核细胞周期调控的主要元件，详细论述见第17章。在脊椎动物细胞中，随着细胞周期进入不同的时期，Cdk酶连续的开启和关闭。当一种这类激酶开启时，通过磷酸化相应的蛋白质来影响细胞行为的各个方面。

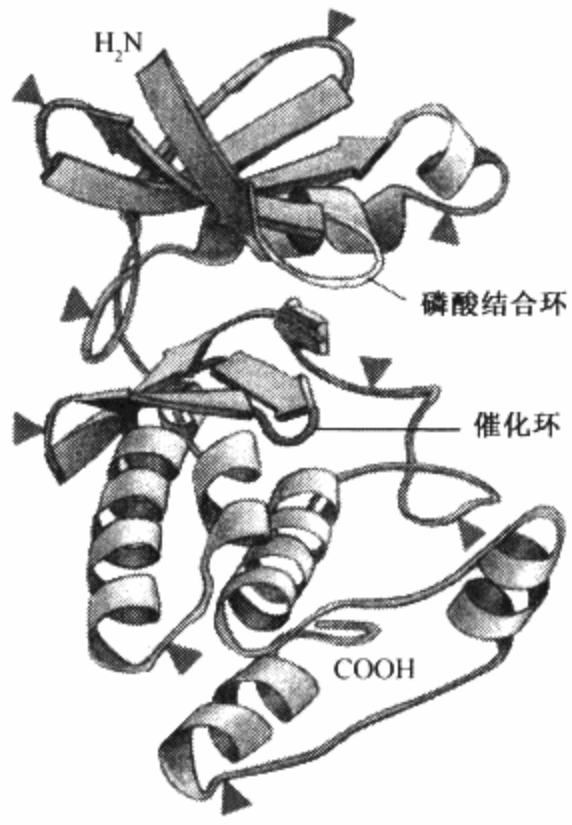


图3-64　蛋白激酶的三维结构

该结构上的双重红色箭头表示在某些蛋白激酶家族中5～100个氨基酸的插入。这个插入位于激酶表面的环区，可以和别的配体相互作用。因此，它们是区分蛋白激酶的标志，可以使其与其他蛋白质相互作用。ATP（贡献磷酸基团）和将被磷酸化的蛋白质位于活性中心，延伸至磷酸结合环区（黄色）和催化环区（橙色）（图3-12）

Cdk蛋白必须和另外一个称为细胞周期蛋白的蛋白质结合才能发挥其丝氨酸/苏氨酸蛋白激酶作用。但是如图3-66所示，与细胞周期蛋白的结合是激活Cdk所需的三个步骤中的一个。除了与细胞周期蛋白结合，磷酸基团必须加到特殊的苏氨酸侧链上，且蛋白质上别的磷酸基团必须去除（共价结合到一个特殊的酪氨酸侧链上）。Cdk控制了一套特异的细胞组分——细胞周期蛋白，一种蛋白激酶和蛋白磷酸酯酶，其作为输入-输出引擎，只有在每个组分都成为合适的活化状态时才会开启。在细胞周期的不同时期，细胞周期蛋白的浓度有高有低，其浓度会逐渐增加直到在周期的一个特定的时刻被突然破坏。细胞周期蛋白的突然降解（通过蛋白水解）会迅速关闭其相应的Cdk酶，而且促进了细胞周期的进程。

类似地，Src蛋白激酶家族也有微芯片的功能。Src蛋白是第一个被发现的酪氨酸激酶，它是只有在多细胞动物中存在的9个相似的蛋白激酶家族中的一个亚家族的成员。根据图3-65的进化树，序列比对指出酪氨酸激酶是一个在进化上较晚的事件，其从丝苏氨酸激酶分支出来的，Src亚家族只是酪氨酸激酶的一小类。

Src蛋白和其同源物包含一个N端区域，它共价连接到一个强疏水性的脂肪酸上，使该激酶在细胞质膜上面向胞内。然后是两个多肽结合结构域，一个Src同源3（SH3）结构域和一个

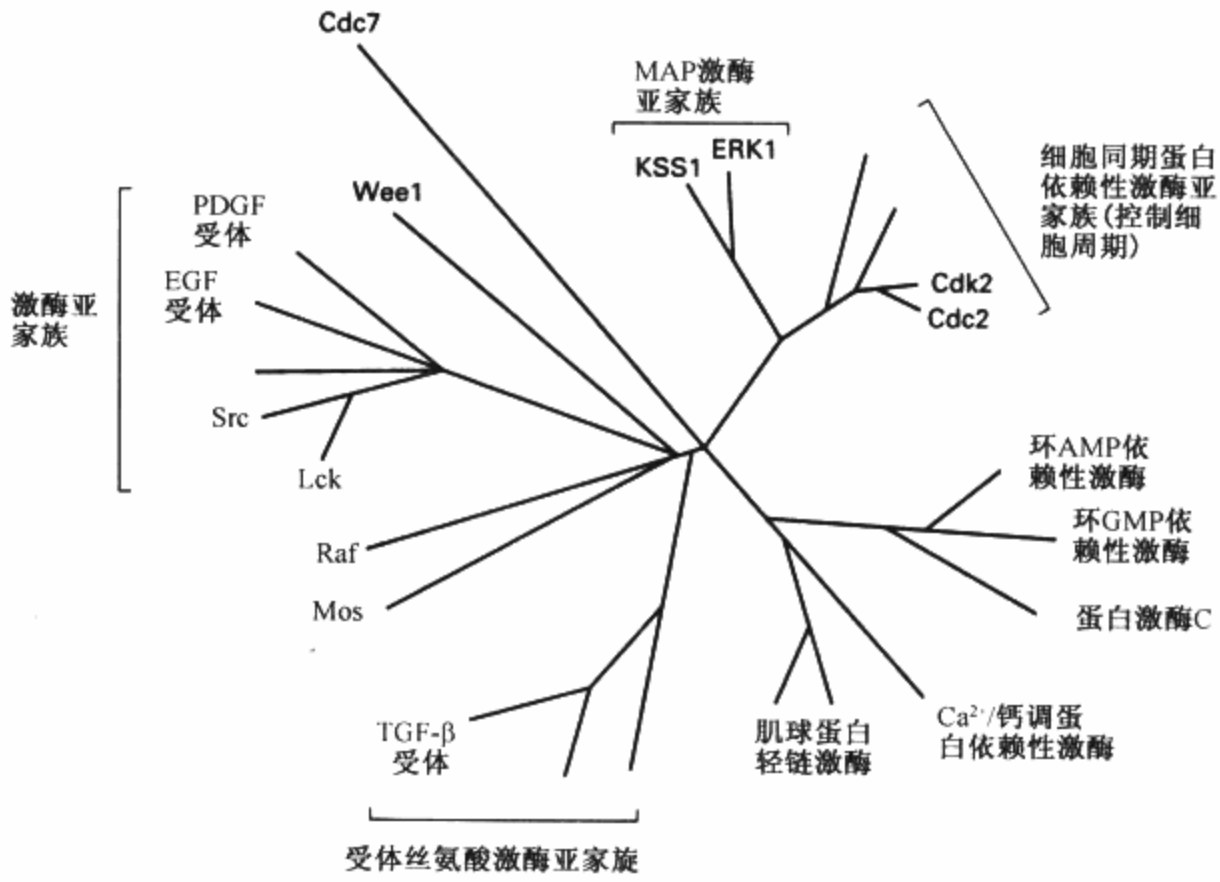


图 3-65　一些蛋白激酶的进化树

虽然高等生物包含有数千种激酶，且人类基因组编码 500 多种此类酶，但本书中只讨论了一部分

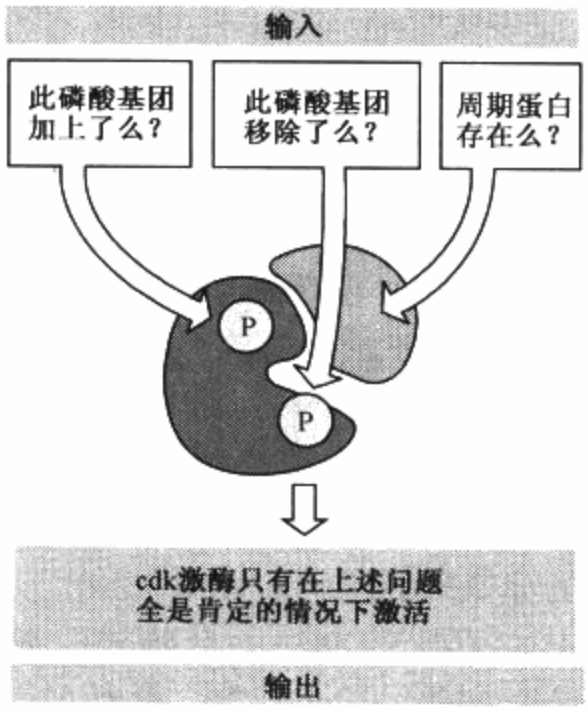


图 3-66　Cdk 蛋白如何作为一个整和装置起作用

它是细胞周期的中心调节因子，其功能见第 17 章

SH2 结构域，其后是一个激酶催化结构域（图 3-67）。这些激酶在正常情况下是处于非活性的状态，C 端的一个磷酸化的酪氨酸结合到 SH2 结构域上，SH3 结构域结合到一个内源肽上，它可以激活该酶的催化位点来帮助解除其无活性状态（图 3-12）。

开启蛋白激酶涉及至少两个步骤：去除 C 端的磷酸基团，通过一个特异的激活蛋白质结合到 SH3 结构域上（图 3-68）。对于 Cdk 蛋白，Src 激酶信号的活化会导致一系列独立的上游事件的完成（图 3-69）。因此，Cdk 和 Src 蛋白家族作为特殊的信号整合因子，帮助产生一系列复杂的信号通路网络，允许细胞根据复杂的环境做出合适的逻辑应答。

蛋白质结合和水解 GTP 是普遍存在的细胞调节因子

我们已经描述了如何通过对蛋白质添加和去除

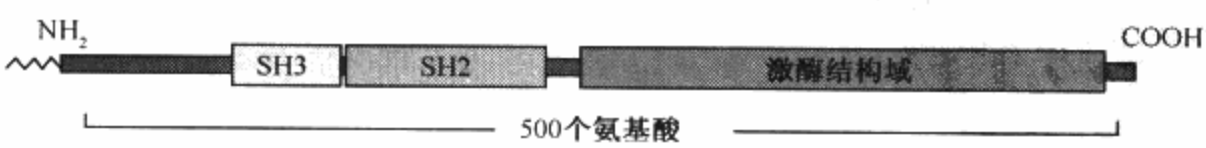


图 3-67　蛋白激酶 Src 家族结构域的结构，根据氨基酸序列作图

磷酸基团来调控蛋白质活性。根据已有的例子，磷酸基团从 ATP 分子上转移到由蛋白激酶催化的蛋白质的氨基酸侧链上。真核细胞也有另一条途径来控制蛋白质加磷酸去磷酸的活性。在这种情况下，磷酸基团不是直接加到蛋白质上；它作为GTP的一部分，GTP 和蛋白质紧密的结合。通常情况下，受这种方式调控的蛋白质在结合 GTP 时是处于有活性的构象的。当结合的 GTP 由蛋白质本身催化水解成 GDP 时，就会解离下磷酸基团，当蛋白质结合 GDP 时是没有活性的。以这种方式，GTP 结合蛋白作为一个开关，其活性由是否有多余的磷酸和 GDP 结合来调控（图 3-70）。

GTP 结合蛋白（也叫做 **GTPase**，因为其催化 GTP 水解）组成了一个大的蛋白质家族，其成员都包含一个结合 GTP 的球状结构域。当紧密结合的 GTP 水解为 GDP 时，这个结构域会发生构象变化而使其失活。图 3-71 中显示了一个单体 GTPase Ras 的三维结构，它是这个家族的一个典型代表。

Ras 蛋白在细胞信号转导中有重要的作用（讨论见第 15 章）。与 GTP 结合时，它是有活性的，而且可以激活下游蛋白磷酸化的级联反应。但是，大多数时间此蛋白质是与 GDP 结合的，没有活性。当应答细胞外信号时，如生长因子，这些信号和细胞质膜上的受体结合，Ras 就会将其 GDP 与 GTP 交换而变成有活性的（图15-55）。

调控蛋白通过决定 GTP 或者 GDP 结合来控制 GTP 结合蛋白的活性

GTP 结合蛋白由调控蛋白控制，它决定是 GTP 还是 GDP 来结合，正如由蛋白激酶和磷酸酯酶来开启或关闭蛋白质的磷酸化。因此，GTPase 活化蛋白（GAP）使 Ras 失活，GAP 结合到 Ras 蛋白上，使 GTP 水解为 GDP（GDP 仍然紧密结合）和无机磷（Pi）（快速地释放）。Ras 蛋白保持其无活性的 GDP 结合构象，直到它遇到一个鸟嘌呤核苷交换因子（GEF），GEF 结合到 GDP-Ras 上，使其释放 GDP。因为空的核苷结合位点迅速被 GTP 分子占领（细胞中 GTP 的量远远高于 GDP），GEF 通过间接加上由 GTP 水解下来的磷酸激活 Ras。因此在某种意义上，GAP 和 GEF 分别与蛋白磷酸酯酶和蛋白激酶类似（图 3-72）。

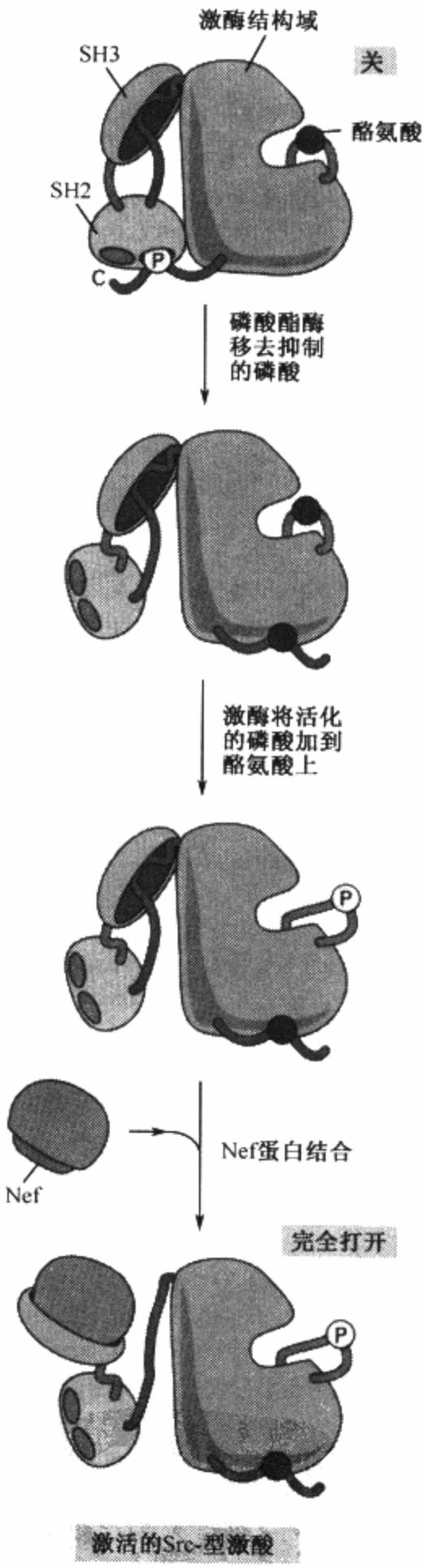


图 3-68　Src 型蛋白激酶通过两步反应激活

大蛋白质的运动来自于小蛋白质

Ras 蛋白属于一大类单体 GTPase 超家族，每个蛋白质包含一个大约由 200 个氨基酸组

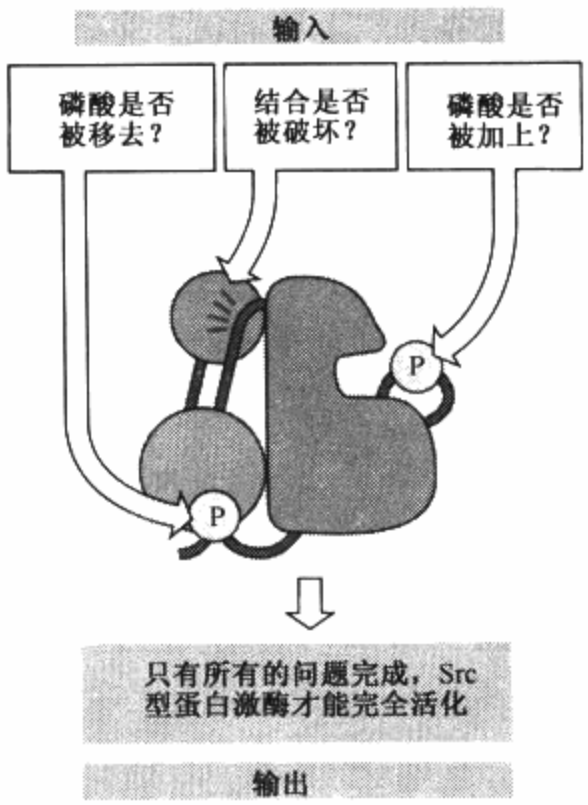


图 3-69　Src 型蛋白激酶一个整合装置起作用

SH3 结构域与 Nef 蛋白的更紧密的相互作用取代其与图中所示的红色连接区域的结合，从而破坏了 SH3 结构域相互作用（绿色）

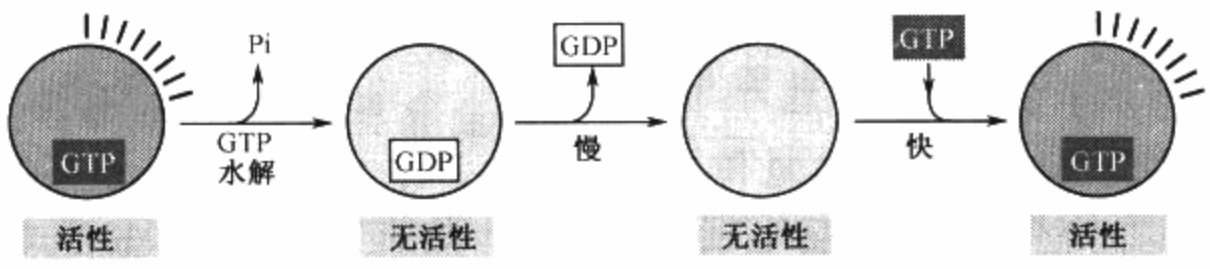


图 3-70　GTP 结合蛋白是分子开关

GTP 结合蛋白（GTPase）的活化通常需要存在一个紧密结合的 GTP 分子（开）。水解该 GTP 产生 GDP 和无机磷（Pi），这使蛋白质变为通常无活性的构象（关）。如图所示，重启开关需要紧密结合的 GDP 的解离，这是一个缓慢的步骤，通常需要被特异的信号来加速；一旦 GDP 解离，1 分子的 GTP 则会快速结合

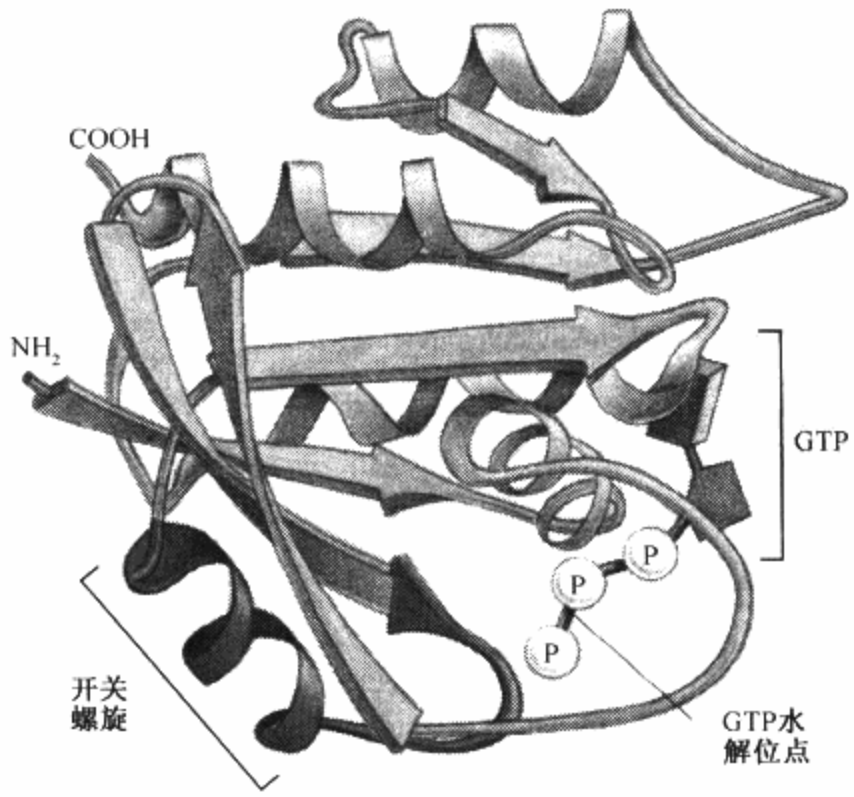


图 3-71　Ras 蛋白结合 GTP 时的构象

单体 GTPase 表明了 GTP 结合结构域的构象，它在一大类 GTP 结合蛋白中都是存在的。当 GTP 分子水解为 GDP 和无机磷时，红色区域的构象会发生变化；GDP 仍然和蛋白质结合，而无机磷被释放。Ras 蛋白的“开关螺旋”的功能将在图 3-74 中描述

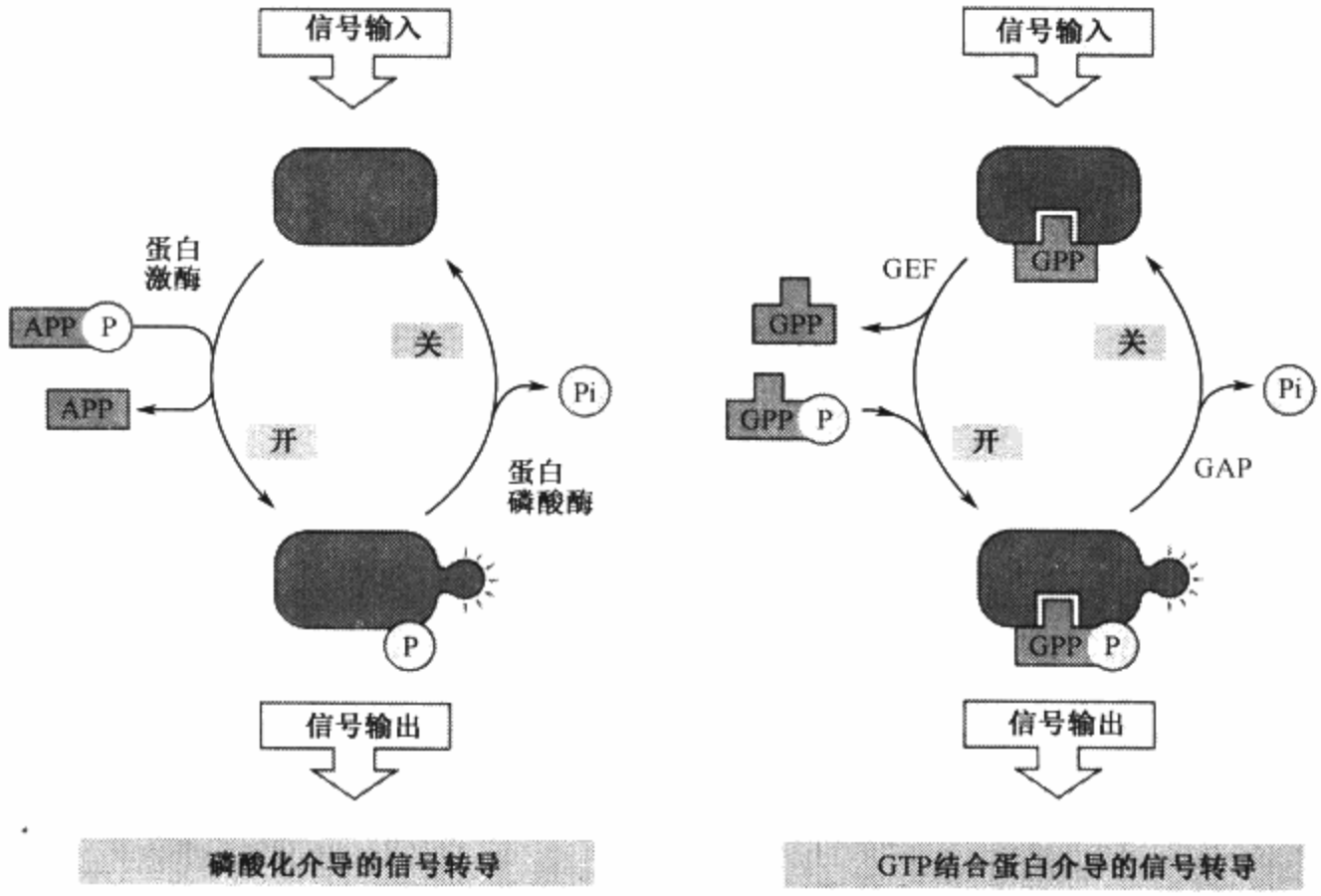


图 3-72　比较真核细胞中两种主要的胞内信号转导机制

在两种情况下，都是通过磷酸基团的加入和去除来激活或者失活信号分子。为了强调两条途径的相似性，ATP 和 GTP 以 APPP 和 GPPP 来表示，而 ADP 和 GDP 分别以 APP 和 GPP 来表示。如图 3-63 所示，蛋白质的磷酸化也可以抑制蛋白质的功能

成的单体 GTP 结合结构域。在进化的过程中，这个结构域可以和那些有多余结构域的大蛋白质结合，形成一大类 GTP 结合蛋白质家族。家族成员包含在细胞信号转导中的受体偶联的三聚体 G 蛋白（讨论见第 15 章），调控胞内细胞器小泡运输的蛋白质（讨论见第 13 章），以及转运 RNA 相结合参与蛋白质合成的组装蛋白质（讨论见第 6 章）。在每种情况下，都是由一个 Ras 类似结构域水解 GTP 改变蛋白质的构象来调控生物反应的活性。

EF-Tu 蛋白是这类家族蛋白质的一个很好的例子。EF-Tu 是一种含量很丰富的分子，它可以作为蛋白质合成中的延伸因子，使每个氨酰 tRNA 分子结合到核糖体上。tRNA 分子和 GTP 结合形式的 EF-Tu 结合形成一个紧密的复合物（图 3-73）。在这个复合物中，结合 tRNA 的氨基酸位点被遮盖。在蛋白质合成时 tRNA 需要暴露其氨基酸结合位点，当与 EF-Tu 结合的 GTP 水解时，其发生在核糖体上，允许 tRNA 解离。因为 GTP 水解是由 tRNA 和 mRNA 在核糖体上有合适的位点，EF-Tu 作为组装因子来区分正确的和不正确的 mRNA-tRNA 配对（图 6-66 有关于 EF-Tu 的详细的功能介绍）。

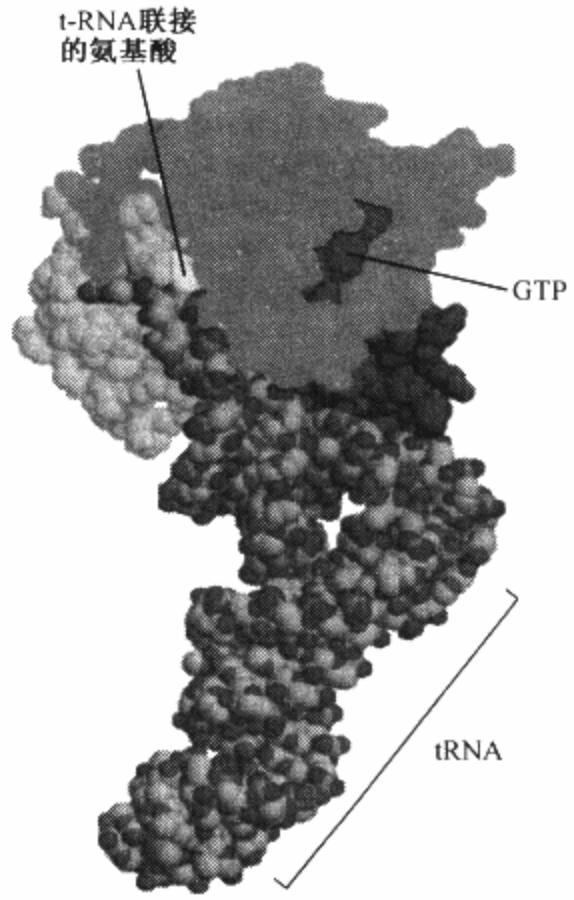


图 3-73　与 EF-Tu 结合的氨酰 tRNA 分子

EF-Tu 蛋白的三个结构域以不同颜色标出，与图 3-74 匹配。该蛋白是细菌蛋白；然而真核生物中存在非常相似的蛋白，此蛋白称为 EF-1。

通过比较 GTP 结合和 GDP 结合时的 EF-Tu 的三维构象可以看出是如何遮盖 tRNA 的。解离的无机磷酸基团（Pi），反应 GTP ⟶GDP＋Pi 的发生使 GTP 的结合位点发生了微小的移动，正如 Ras 蛋白一样。这个小的移动，与氢原子的直径的数量级相当，使构象发生变化而使 α 螺旋向前扩散，叫做螺旋转换，这发生在蛋白质的 Ras 类似结构域。转换螺旋作为门插销粘到分子别的结构域上的一个特异位点，使蛋白质处于一个关闭的状态。由 GTP 水解催化的构象变化使转换螺旋分开，使蛋白质的各个结构域摆动分开，通过一个约 4nm 的距离。这样释放结合的 tRNA 分子，使结合的氨基酸可以被利用（图 3-74）。

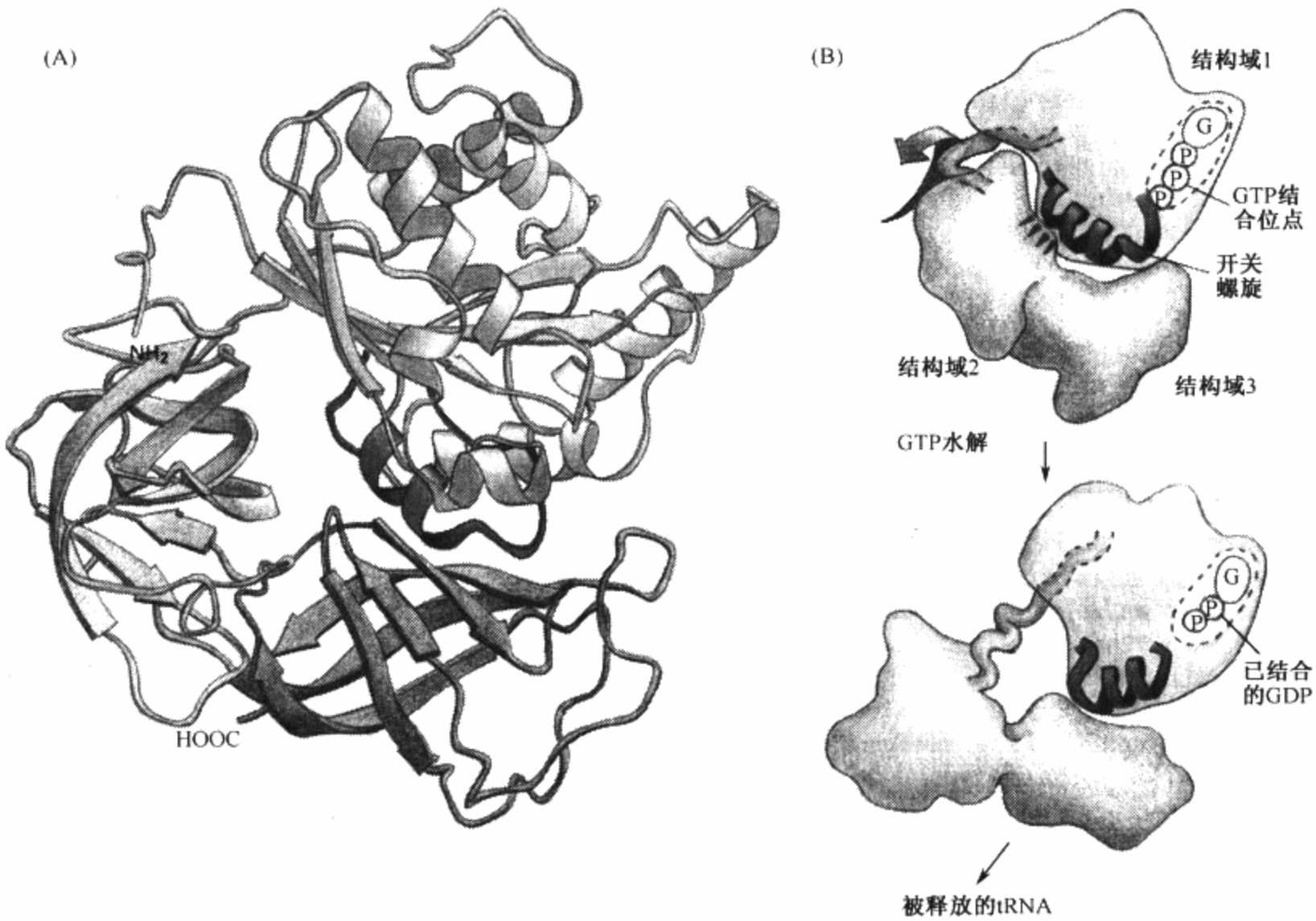


图 3-74　GTP 水解引起的 EF-Tu 大规模构象改变

（A）结合 GTP 的 EF-Tu 的三维结构，顶端结构域和 Ras 蛋白同源，红色的 α-螺旋是开关螺旋，可以在 GTP 水解后移动，如 3-71 所示（B）结构域 I 中开关螺旋构型的改变使得结构域 2 和 3 作为单个基团向读者方向旋转 90°，继而释放图 3-73 中所示的结合其上的 tRNA.［（A）摘自 H. Berchtole et al.，*Nature* 365：126～132，1993；（B）经 Mathisa Sprinzl 和 Rolf Hilgenfeld 许可］

从这个例子中我们可以看出细胞怎样利用发生在蛋白质结构域表面简单的化学变化来创造一个 50 倍的运动。这种巨大的结构变化也发生在我们接下来将会讲解的动力蛋白的运动中。

动力蛋白负责细胞中的大运动

已经知道蛋白质的构象变化是怎样在酶调节和细胞信号传递的过程中发挥核心作用。现在我们来讨论那些主要功能是移动其他分子的蛋白质。这些**动力蛋白**产生了肌肉收缩和细胞爬行游动的力。动力蛋白还给小范围的细胞内运动提供了动力：它们帮助有丝分裂中染色体向细胞两极移动（讨论见第 18 章），帮助细胞器沿着细胞内的分子轨道移动（讨论见第 16 章），在合成新的 DNA 分子时帮助酶在 DNA 链上移动（讨论见第 5 章）。所有这些基础过程

都依赖于那些具有活动部分的蛋白质，这些蛋白质可以作为产生动力的机器。

这些机器怎样工作呢？换句话说，这些蛋白质的形状变化是怎样产生细胞内的直接运动呢？例如，如果一个蛋白质需要沿着一条窄线（如 DNA 分子）走动，它可以通过一系列的构象变化来实现这一过程，如图 3-75 所示。但是，如果没有东西驱动这些可逆的有序变化，这些蛋白质就可以在线上前后徘徊。我们可以用另一方式来解释这一情况。因为一个蛋白质发生了定向的运动，热力学规律（讨论见第 2 章）要求这种运动利用源自其他东西的自由能（否则这种蛋白质就可以用来制造永动机了）。因此，没有能量的输入，蛋白质分子只能无目标地徘徊。

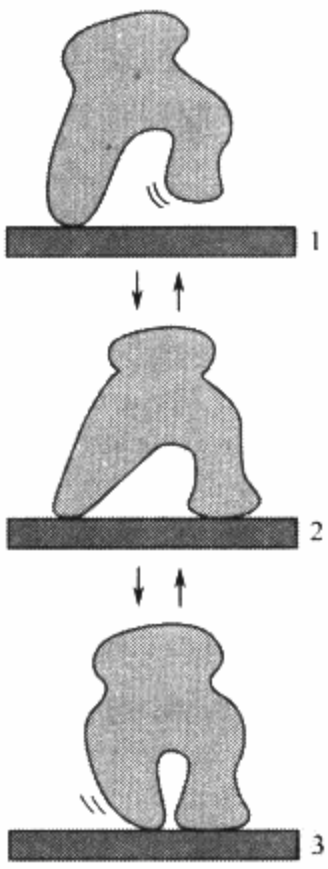


图 3-75　一个变构“走动”的蛋白质

尽管在与一条线或丝结合时，它的三种构象允许它随机地前后徘徊，但这个蛋白质不能朝着一个单一的方向移动

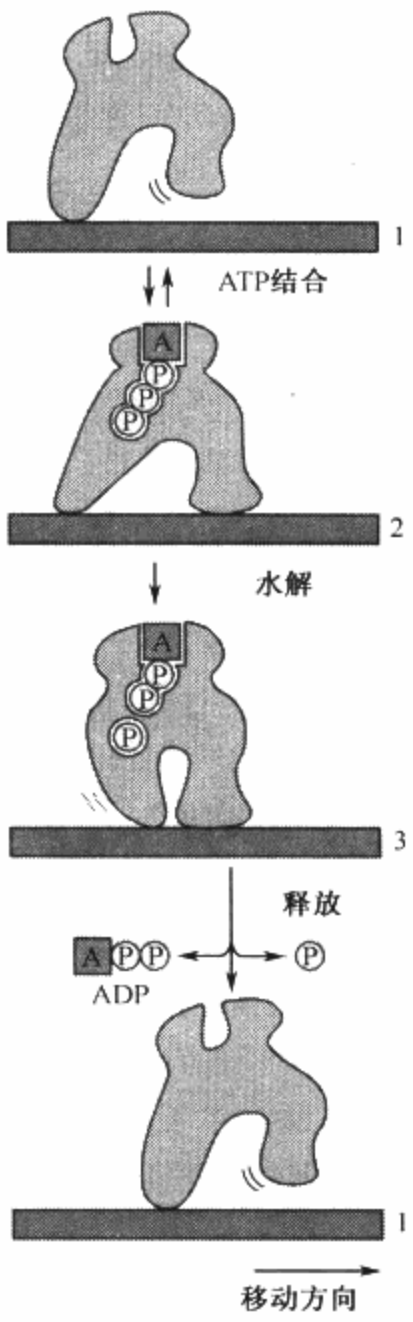


图 3-76　一个变构的动力蛋白

三种不同构象的转变包括由结合的 ATP 分子水解驱动的一步，这使得这个循环不可逆。通过重复这种循环，蛋白质沿着线向右方不断移动

那么，一个蛋白质怎样使一系列的构象变化向着单一的方向变化呢？为了使整个循环向一个方向进行，只要让任何一种形状改变成为不可逆的就足够了。对于大多数能够沿着一个方向长距离移动的蛋白质来说，通过将一种构象变化和与蛋白质结合的 ATP 分子水解相偶联来实现这一过程的。这种机制类似于刚刚讨论过的驱动变构蛋白形状改变的 GTP 水解过程。由于 ATP（或 GTP）水解时释放大量自由能，核苷酸结合蛋白质进行向后移动的可逆形状变

化就不太可能了，因为这需要逆转 ATP 的水解过程，即在 ADP 上添加一个磷酸基团形成 ATP。

在图 3-76 显示的模型中，ATP 的结合使一个动力蛋白从构象 1 转变到构象 2。然后，结合的 ATP 分子水解产生 ADP 和无机磷酸（Pi），引发从构象 2 到构象 3 的变化。最后，ADP 和 Pi 的释放驱动蛋白质回到构象 1。由于构象 2 到构象 3 的转变是由 ATP 水解提供的能量驱动的，所以这一系列的构象变化实际上是不可逆的。因此，整个循环只向一个方向前进，在这个例子中使得蛋白质分子不停地走向右方。

许多动力蛋白通过这种普遍的方式产生定向的移动，包括肌肉的动力蛋白、肌球蛋白（它可以沿着肌动蛋白丝移动而产生肌肉收缩）和沿着微管走动的驱动蛋白（两种蛋白质都将在第 16 章讨论）。这些运动可以很迅速：一些和 DNA 复制相关的动力蛋白（DNA 解旋酶）以每秒 1000 个核苷酸的速度沿着 DNA 链自我推进。

膜结合转运蛋白利用能量将分子输送过膜

我们已经知道了变构蛋白怎样以微芯片（Cdk 和 Src 激酶）、组装因子（EF-Tu）以及机械力和运动的产生者（动力蛋白）等方式发挥作用。此外，变构蛋白还可以利用来自 ATP 水解的能量、离子梯度或电子传递过程来输送特异离子或小分子跨过膜。我们在这里讨论一个例子，其他的会在第 11 章讨论。

肌肉细胞中的钙离子转运蛋白是了解得最清楚的泵蛋白之一。这种蛋白质叫做**钙泵**，被包埋在肌肉细胞中的特殊细胞器肌质网的膜中。钙泵（也叫钙离子 ATP 酶）通过将钙离子从细胞质泵入肌质网的过程，维持静息肌肉细胞质的低钙离子浓度；然后，响应一个神经刺激时，钙离子通过它们的通道被释放到细胞质中，引发肌肉收缩。钙离子泵与维持细胞质两侧钠离子和钾离子浓度差异的钠离子-钾离子泵同源，它们都是 P 类型的阳离子泵（这一命名是因为这些蛋白质可以在反应周期中自动磷酸化）。一个大的细胞质内的头部区域结合并水解 ATP，与释放的磷酸基团形成共价键。因此蛋白质就短暂的转变为高能的磷酸化的状态，与从细胞质中获得的两个钙离子紧密结合。然后，这种形式的蛋白质能量衰减变成一个低能的磷酸化状态，使钙离子释放到肌质网内腔。最后，酶的去磷酸化释放磷酸分子，使蛋白质开始下一个泵入离子的周期。

所有 P 类型的离子泵的头部区域都是与一系列的 10 个跨膜 α 螺旋连接，其中 4 个形成钙泵的跨膜钙离子结合位点。已经利用高分辨率的电子显微镜和 X 射线衍射解出了这种蛋白质的三维结构（图 11-15）。这就使生物学家们可以基于其正常和突变形式的大量生物化学数据来推演出一种泵活动的分子模型。

如图 3-77（A）中解释的，与钙离子结合的跨膜 α 螺旋互相缠绕，螺旋间形成钙离子通过的孔穴。ATP 水解使在细胞质内的头部区域产生了一系列的重大构象变化，头部区域通过与跨膜螺旋相连的柄改变膜内一些螺旋的结构和相对位置，从而改变孔穴大小，使得钙离子单向地通过膜［图 3-77（B）］。与动力蛋白一样，离子的单向转运需要一个利用能量的循环，来使蛋白质的构象变化具有确定的方向性。

人们发明了很多不同种类的机械泵，所以细胞中同时有以另一方式作用的膜结合的泵也不值得奇怪。这其中最著名的是偶联 ATP 水解和氢离子（质子）转运的旋转泵。这些泵类似微型的涡轮，用于提高溶酶体和其他真核细胞器的酸性。与其他产生离子梯度的离子泵一样，如果它们所转运的离子有足够高的跨膜梯度，也可以反过来催化 ADP 和磷酸生产 ATP 的反应。

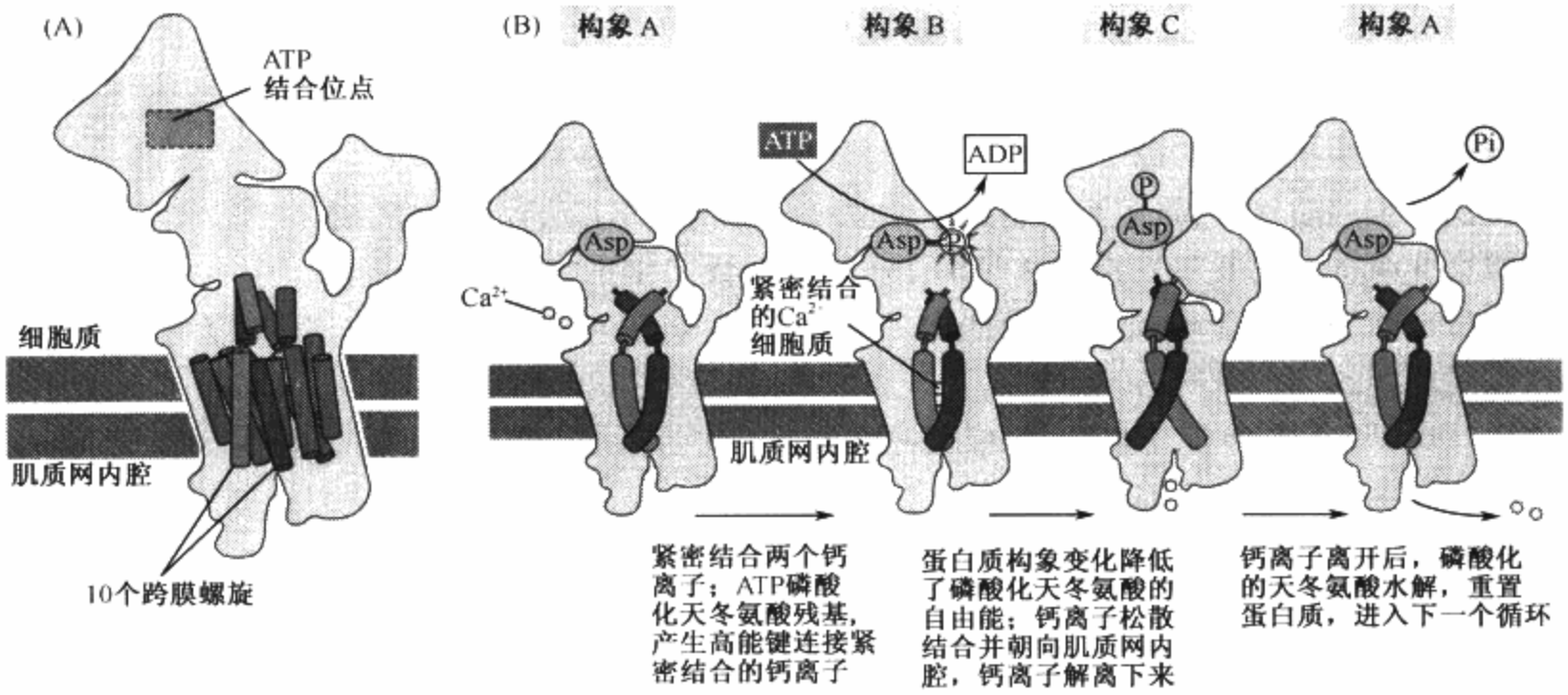


图 3-77　通过钙泵转运钙离子

（A）泵蛋白的结构，由 994 个氨基酸的单亚基构成（更多细节见图 11-15）。（B）离子泵活动的模型。为了简单起见，只显示了 4 个与钙离子结合的跨膜 α 螺旋中的 2 个

一个这样的泵，ATP 合酶，利用电子转运产生的质子浓度梯度来产生生命世界所利用的大多数 ATP。由于这种普遍存在的泵在能量转换的过程中具有重要的核心作用，所以我们将在第 14 章讨论它的三维结构和机制。

蛋白质通常形成巨大的复合体，以蛋白质机器的形式发挥功能

由单结构域的小蛋白质形成多结构域的大蛋白质，蛋白质可以发挥更精细的功能。但是，最重大的任务是由许多蛋白质分子形成的大蛋白质集合体完成的。现在，大多数生命进程都可以在实验室的非细胞体系中重建，已经清楚地知道细胞中的每个核心进程，如 DNA 复制、蛋白质合成、小泡芽生、跨膜的信号传递，都是由高度配合、相互关联的 10 个或更多个蛋白质组成的装置催化的。在大多数的这种蛋白质机器中，结合的核苷三磷酸（ATP 或 GTP）的水解驱动了某些蛋白质亚基的一系列有序的构象变化，使蛋白质组合体可以一起移动。通过这种方法，每种合适的酶都会直接被放到所需要的位置，来进行一系列的连续反应。例如，这一过程在核糖体中的蛋白质合成（讨论见第 6 章）或者 DNA 复制（此过程中一个大的多种蛋白质的复合体会迅速沿着 DNA 移动，讨论见第 5 章）时都有发生。

出于与人类发明机械和电力机器一样的原因，细胞也改进了蛋白质机器。为了完成几乎任何一种任务，通过相关进程在空间上暂时协调的多聚蛋白质比连续使用单一的工具更加有效。

细胞功能的基础是复杂的蛋白质相互作用网络

在已知完整的基因组序列的后基因组时代，细胞生物学家面临着很多挑战，其中一个就是解剖和重建生物体中存在的数千种蛋白质机器。为了理解这些非凡的蛋白质复合物，每一个都必须用其纯化的蛋白质组分来重建，以便于在不含其他细胞组分的试管中研究它在控制条件下的详细作用模式。这一过程本身就是一个巨大的任务。但是我们现在已经知道这些亚

细胞组分中的每一个都要与其他形式的大分子相互作用，在整个细胞中产生一个巨大的蛋白质-蛋白质和蛋白质-核酸相互作用的网络。因此，为了了解细胞，还需要分析这些相互作用的大部分。有关细胞内蛋白网络的复杂性的观念可以从第 16 章描述的一个特殊的已经很好研究过的例子中得到：酵母中许多蛋白质与肌动蛋白（图 16-15）。

这种蛋白质-蛋白质相互作用的程度也可以用一般的方法评估。特别是，利用第 8 章描述的双杂交筛选（图 8-51）检测这些相互作用的大量工作正在进行中。例如，这一方法正应用于酵母产生的一大套 6000 个基因产物。像预期的一样，观测到的相互作用中的大多数都发生在相同功能的蛋白质群内。也就是说，和细胞周期调控相关的蛋白质倾向于互相广泛的相互作用，对于 DNA 合成或者 DNA 修复等过程的相关蛋白质也是如此。但是还有惊人数量的相互作用发生在不同功能的蛋白质群体之间（图 3-78）。这些相互作用大概对协调细胞功能是很重要的，但是大多数这种相互作用还不了解。

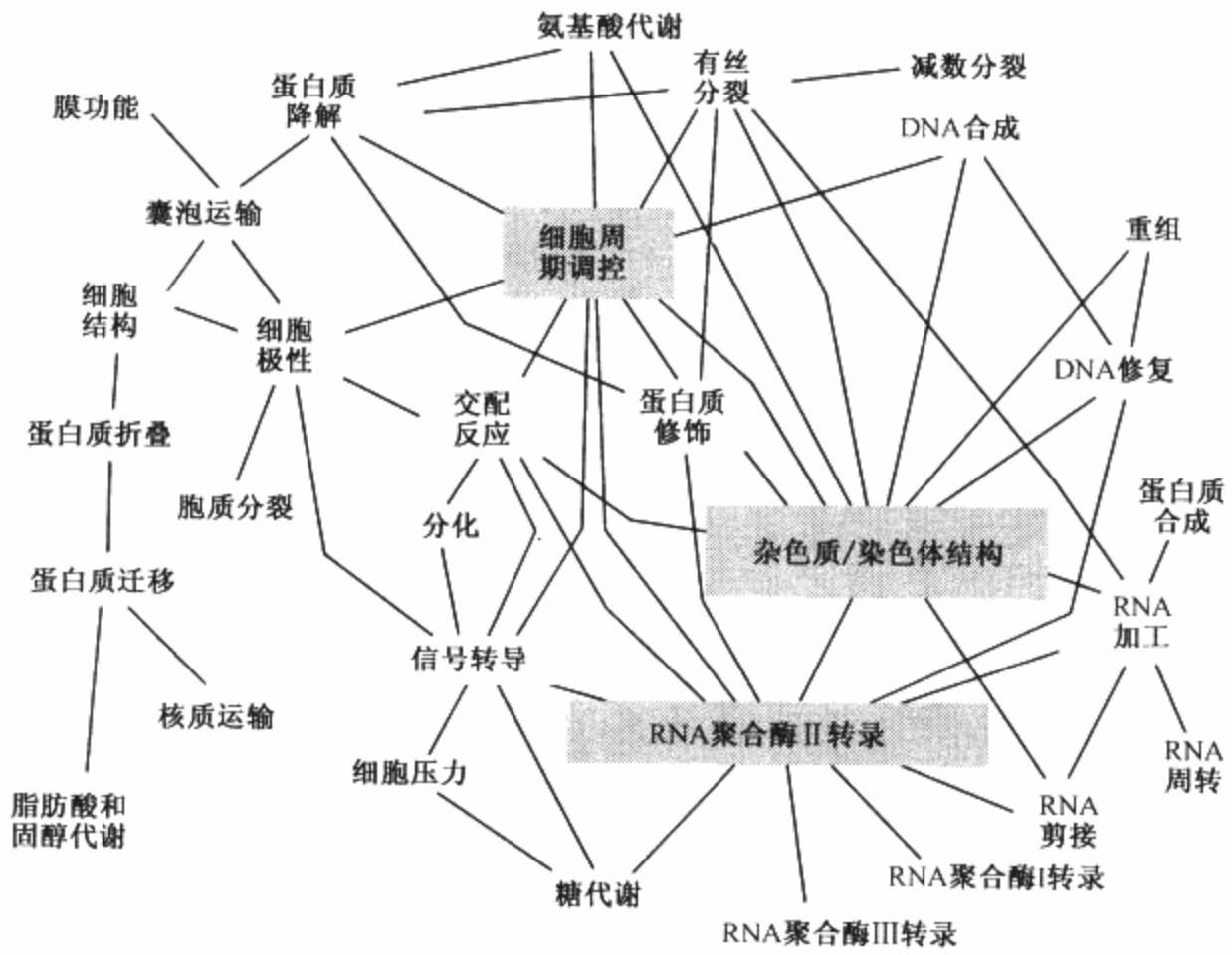


图 3-78　观察到的酵母中不同功能蛋白质群体之间蛋白质-蛋白质相互作用的网络结构图

为了得到这张图，多于 1500 种蛋白质被归为 31 种功能群体。观测到的大约 70%的已知蛋白质-蛋白质相互作用发生在归属于同一功能群体的蛋白质之间。但是，像推测的一样，还有惊人数量的相互作用发生在不同功能的蛋白质群体之间。这张图表的每一条线表明，在被线连接的两个功能群体之间观察到了多于 15 种蛋白质-蛋白质相互作用。三个用黄色高亮显示的功能群体，与 10 个或更多的功能群体之间存在至少 15 种相互作用（摘自 C. L. Tucker, J. F. Gera, and P. Vetz, *Trends Cell Biol.* 11: 102～106, 2001）

对可用数据知识的考查表明，人类细胞中的一般蛋白质可能在某种程度上与 5～15 种不同的伙伴相互作用。通常，不同系列的伙伴会与多结构域蛋白质的不同结构域结合；实际上，可以推测观察到的人类蛋白质的不寻常的大量多结构域的结构产生了这些相互作用。细胞中大分子的相互作用网络的极端复杂，要完全解释它们的功能可能需要科学家们忙碌几个世纪。

小结

蛋白质可以形成极为复杂的化学装置，它们的主要功能大部分决定于它们的化学性质。表面的孔穴形成了配基的结合位点，在这些位置通过蛋白质折叠将精确定位的氨基酸侧链放到一起。利用同样的方法可以激活通常处于非活性状态的氨基酸侧链，来产生或者打断共价键。酶是一种具有催化功能的蛋白质，可以与高能的过渡态结合，通过特殊的路径进行反应，从而大大提高了反应速度；它们也可以同时进行酸催化和碱催化。酶反应的速度常常太快了，以至于它们只受扩散的限制；如果将对一个底物连续作用的酶互相结合形成单一的多酶复合物，或者将酶及其底物限制在细胞内的同一区室，就可以进一步加快反应的速度。

当配基结合到蛋白质表面时，蛋白质可逆地改变形状。这种由于一种配基结合产生的蛋白质构象变化会影响第二种配基的结合，两个配基结合位点之间的联系是调节细胞进程的核心机制。例如，通过反馈调节控制代谢途径：在途径早期，一些小分子起抑制酶的作用而另一些小分子起激活酶的作用。用这一方法控制的酶常常形成对称的集合体，可以引起协调的构象变化来对调节它们作用的配体浓度变化做出快速反应。

蛋白质形状变化是由消耗化学能的单向过程驱动的。例如，通过将形状变化和 ATP 水解偶联在一起，蛋白质可以有用的工作，如产生向单一方向长距离移动的机械力。由 X 射线测定的蛋白质三维结构，已经揭示了核苷三磷酸水解产生的局部小改变怎样扩大到产生蛋白质其他部位的大改变。通过这种方法，这些蛋白质可以作为信息传递的输入-输出装置，如组装因子、动力蛋白、膜结合泵。将许多不同的蛋白质分子结合到一起形成了高效的蛋白质机器，机器的各个组分的变构运动都是协调进行的。现在我们已经知道，这样的机器进行了细胞内的许多最为重要的反应。

参考文献

一般文献

Branden C & Tooze J (1999) Introduction to Protein Structure, 2nd edn. New York: Garland Publishing.

Creighton TE (1993) Proteins: Structures and Molecular Properties, 2nd edn. New York: WH Freeman.

Dickerson RE & Geis I (1969) The Structure and Action of Proteins. New York: Harper & Row.

Kyte J (1995) Structure in Protein Chemistry. New York: Garland Publishing.

Mathews CK, van Holde KE & Ahern K-G (2000) Biochemistry, 3rd edn. San Francisco: Benjamin Cummings.

Perutz M (1992) Protein Structure: New Approaches to Disease and Therapy. New York: WH Freeman.

Schulz GE & Schirmer RH (1990) Principles of Protein Structure, 2nd edn. New York: Springer.

Stryer L (1995) Biochemistry, 4th edn. New York: WH Freeman.

The Shape and Structure of Proteins

Anfinsen CB (1973) Principles that govern the folding of protein chains. *Science* 181, 223–230.

Bowie JU & Eisenberg D (1993) Inverted protein structure prediction. *Curr. Opin. Struct. Biol.* 3, 437–444.

Burkhard P, Stetefeld J & Strelkov SV (2001) Coiled coils: a highly versatile protein folding motif. *Trends Cell Biol.* 11, 82–88.

Caspar DLD & Klug A (1962) Physical principles in the construction of regular viruses. *Cold Spring Harb. Symp. Quant. Biol.* 27, 1–24.

Doolittle RF (1995) The multiplicity of domains in proteins. *Annu. Rev. Biochem.* 64, 287–314.

Fraenkel-Conrat H & Williams RC (1955) Reconstitution of active tobacco mosaic virus from its inactive protein and nucleic acid components. *Proc. Natl Acad. Sci. USA* 41, 690–698.

Fuchs E (1995) Keratins and the skin. *Annu. Rev. Cell Dev. Biol.* 11, 123–153.

Gehring WJ, Affolter M & Burglin T (1994) Homeodomain proteins. *Annu. Rev. Biochem.* 63, 487–526.

Harrison SC (1992) Viruses. *Curr. Opin. Struct. Biol.* 2, 293–299.

International Human Genome Sequencing Consortium (2001) Initial sequencing and analysis of the human genome. *Nature* 409, 860–921.

Marchler-Bauer A & Bryant SH (1997) A measure of success in fold recognition. *Trends Biochem. Sci.* 22, 236–240.

Nomura M (1973) Assembly of bacterial ribosomes. *Science* 179, 864–873.

Pauling L & Corey RB (1951) Configurations of polypeptide chains with favored orientations around single bonds: two new pleated sheets. *Proc. Natl. Acad. Sci. USA* 37, 729–740.

Pauling L, Corey RB & Branson HR (1951) The structure of proteins: two hydrogen-bonded helical configurations of the polypeptide chain. *Proc. Natl. Acad. Sci. USA* 27, 205–211.

Pawson T (1994) SH2 and SH3 domains in signal transduction. *Adv. Cancer Res.* 64, 87–110.

Ponting CP & Dickens NJ (2001) Genome cartography through domain annotation. *Genome Biol.* 2(7), comment2006.1–2006.7.

Ponting CP, Schultz J, Copley RR et al. (2000) Evolution of domain families. *Adv. Protein Chem.* 54, 185–244.

Richards FM (1991) The protein folding problem. *Sci. Am.* 264(1), 54–63.

Richardson JS (1981) The anatomy and taxonomy of protein structure. *Adv. Protein Chem.* 34, 167–339.

Sali A & Kuriyan J (1999) Challenges at the frontiers of structural biology. *Trends Cell Biol.* 9, M20–M24.

Steiner DF, Kemmler W, Tager HS & Peterson JD (1974) Proteolytic processing in the biosynthesis of insulin and other proteins. *Fed. Proc.* 33, 2105–2115.

Teichmann SA, Murzin AG & Chothia C (2001) Determination of protein function, evolution and interactions by structural genomics. *Curr. Opin. Struct. Biol.* 11, 354–363.

Trinick J (1992) Understanding the functions of titin and nebulin. *FEBS Lett.* 307, 44–48.

Protein Function

Alberts B (1998) The cell as a collection of protein machines: preparing the next generation of molecular biologists. *Cell* 92, 291–294.

Benkovic SJ (1992) Catalytic antibodies. *Annu. Rev. Biochem.* 61, 29–54.

Berg OG & von Hippel PH (1985) Diffusion-controlled macromolecular interactions. *Annu. Rev. Biophys. Biophys. Chem.* 14, 131–160.

Bourne HR (1995) GTPases: a family of molecular switches and clocks. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 349, 283–289.

Braden BC & Poljak RJ (1995) Structural features of the reactions between antibodies and protein antigens. *FASEB J.* 9, 9–16.

Dickerson RE & Geis I (1983) Hemoglobin: Structure, Function, Evolution and Pathology. Menlo Park, CA: Benjamin Cummings.

Dressler D & Potter H (1991) Discovering Enzymes. New York: Scientific American Library.

Fersht AR (1999) Structure and Mechanisms in Protein Science: A Guide to Enzyme Catalysis. New York: WH Freeman.

Hazbun TR & Fields S (2001) Networking proteins in yeast. *Proc. Natl. Acad. Sci. USA* 98, 4277–4278.

Hunter T (1987) A thousand and one protein kinases. *Cell* 50, 823–829.

Jones S & Thornton JM (1996) Principles of protein–protein interactions. *Proc. Natl Acad. Sci. USA* 93, 13–20.

Kantrowitz ER & Lipscomb WN (1988) *Escherichia coli* aspartate transcarbamoylase: the relation between structure and function. *Science* 241, 669–674.

Khosla C & Harbury PB (2001) Modular enzymes. *Nature* 409, 247–252.

Koshland DE, Jr (1984) Control of enzyme activity and metabolic pathways. *Trends Biochem. Sci.* 9, 155–159.

Kraut J (1977) Serine proteases: structure and mechanism of catalysis. *Annu. Rev. Biochem.* 46, 331–358.

Lichtarge O, Bourne HR & Cohen FE (1996) An evolutionary trace method defines binding surfaces common to protein families. *J. Mol. Biol.* 257, 342–358.

Marcotte EM, Pellegrini M, Ng HL et al. (1999) Detecting protein function and protein–protein interactions from genome sequences. *Science* 285, 751–753.

Miles EW, Rhee S & Davies DR (1999) The molecular basis of substrate channeling. *J. Biol. Chem.* 274, 12193–12196.

Monod J, Changeux J-P & Jacob F (1963) Allosteric proteins and cellular control systems. *J. Mol. Biol.* 6, 306–329.

Pavletich NP (1999) Mechanisms of cyclin-dependent kinase regulation: structures of Cdks, their cyclin activators, and Cip and INK4 inhibitors. *J. Mol. Biol.* 287, 821–828.

Perutz M (1990) Mechanisms of Cooperativity and Allosteric Regulation in Proteins. Cambridge: Cambridge University Press.

Phillips DC (1967) The hen egg white lysozyme molecule. *Proc. Natl. Acad. Sci. USA* 57, 484–495.

Radzicka A & Wolfenden R (1995) A proficient enzyme. *Science* 267, 90–93.

Schramm VL (1998) Enzymatic transition states and transition state analog design. *Annu. Rev. Biochem.* 67, 693–720.

Schultz PG & Lerner RA (1995) From molecular diversity to catalysis: lessons from the immune system. *Science* 269, 1835–1842.

Sicheri F & Kuriyan J (1997) Structures of Src-family tyrosine kinases. *Curr. Opin. Struct. Biol.* 7, 777–785.

Todd AE, Orengo CA & Thornton JM (1999) Evolution of protein function, from a structural perspective. *Curr. Opin. Chem. Biol.* 3, 548–556.

Toyoshima C, Nakasako M, Nomura H & Ogawa H (2000) Crystal structure of the calcium pump of sarcoplasmic reticulum at 2.6Å resolution. *Nature* 405, 647–655.

Vale RD & Milligan RA (2000) The way things move: looking under the hood of molecular motor proteins. *Science* 288, 88–95.

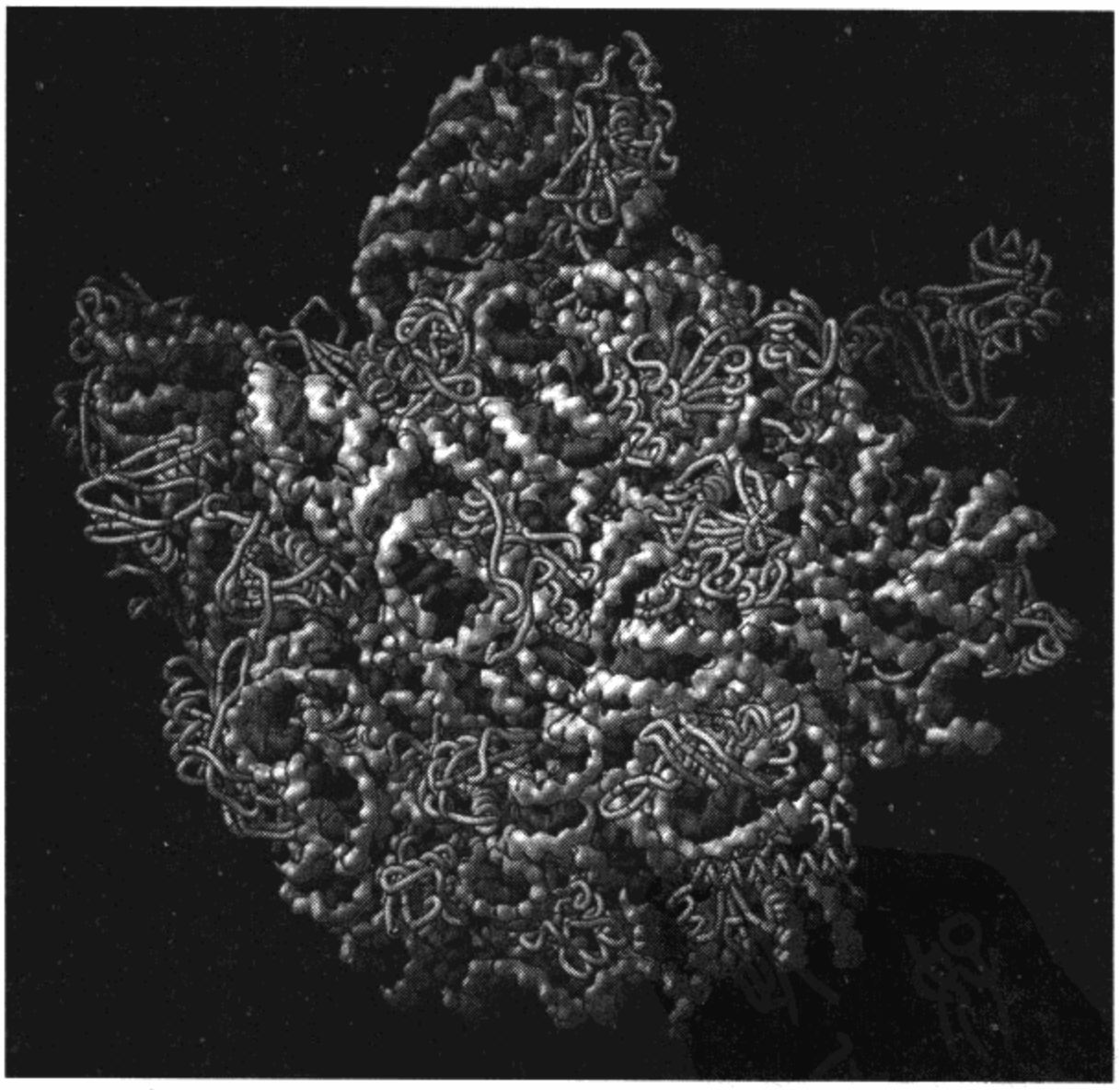
Vocadlo DJ, Davies GJ, Laine R & Withers SG (2001) Catalysis by hen egg-white lysozyme proceeds via a covalent intermediate. *Nature* 412, 835–838.

Walsh C (2001) Enabling the chemistry of life. *Nature* 409, 226–231.

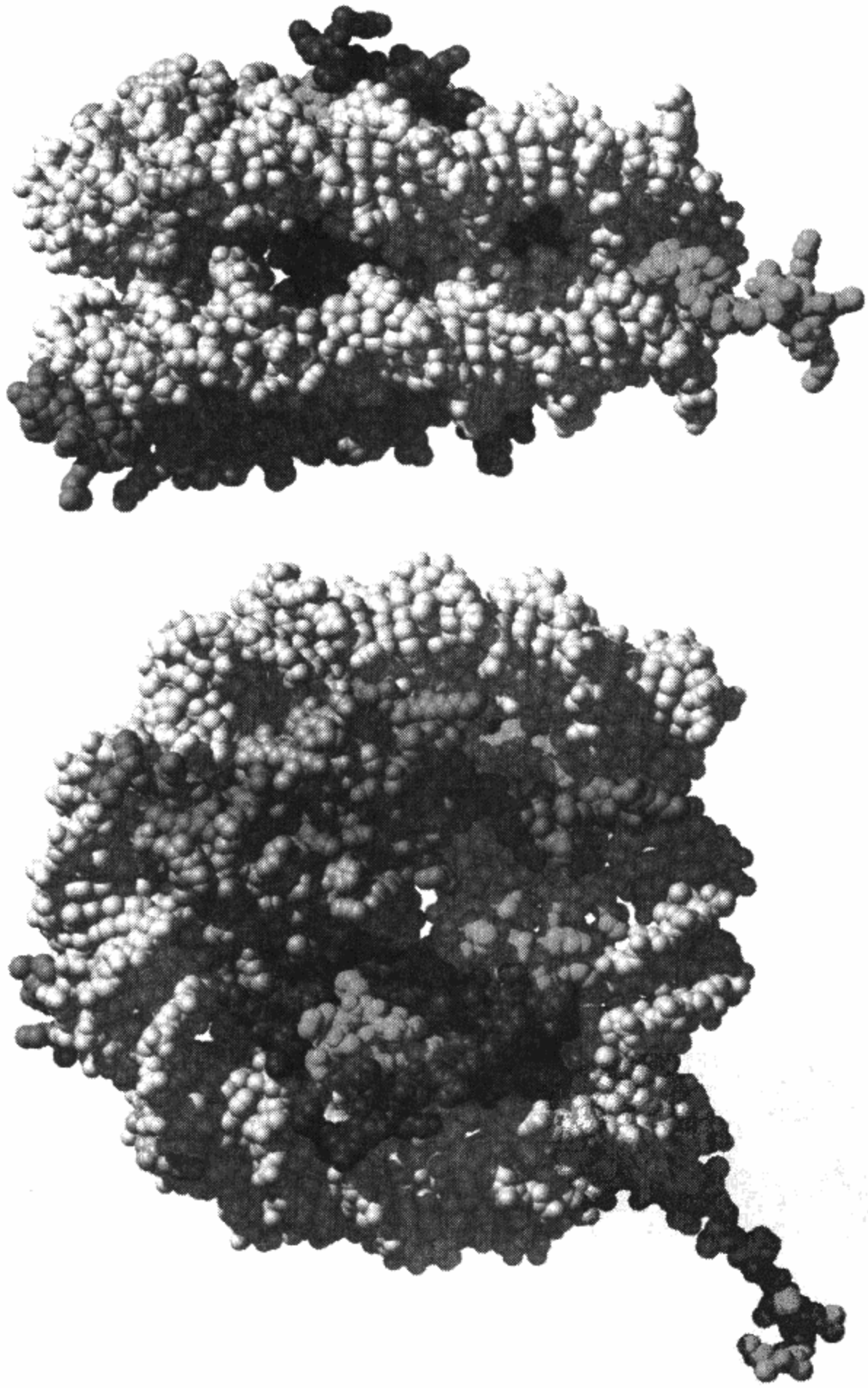
Wolfenden R (1972) Analog approaches to the structure of the transition state in enzyme reactions. *Acc. Chem. Res.* 5, 10–18.

第 II 部分　基本遗传机制

4　DNA 与染色体
5　DNA 复制、修复以及重组
6　细胞如何解读基因组
7　基因表达的调控



核糖体。图示为细菌核糖体大亚基，其中核糖体 RNA 用灰色标出，核糖体蛋白用金黄色标出。核糖体是蛋白质合成中处于核心地位的复杂合成机器（摘自 N. Ban et al.，*Science* 289：905～920，2000. © AAAS.）



核小体。所有真核细胞染色体的基本结构单元是核小体。DNA 双螺旋（灰色）被组蛋白核心颗粒彩色包围，形成核小体。核小体在染色体 DNA 上约间隔 200 个核苷酸对分布（许可翻印 K. Luger et al.，*Nature* 389：251～260，1997. © Macmillan Magazines Ltd.）

4 DNA 与染色体

DNA 的结构和功能

染色体 DNA 及其在染色质纤维中的包装

染色体的总体结构

生命的存在依赖于细胞存储、检索，并解译形成和维持活生物体所需遗传指令的能力。这种**遗传**信息在细胞分裂过程中从一个细胞传递到它的子细胞，并通过生物体的繁殖细胞从上一代传递到下一代。每个活细胞内都储存了这些指令作为它的**基因**，即含有信息的基本单位，它不仅能够决定某一物种的特征，也能决定这一物种内个体的特征。

20 世纪初，遗传学刚开始成为一门学科的时候，科学家们就对基因的化学结构产生了浓厚的兴趣。在多细胞生物的生命过程中，储存在基因内的信息经过拷贝，并从细胞传递到子细胞中达数百万次，从本质上讲，这些信息在此过程中没有发生改变。什么样的分子能够经历如此精确和几乎无限次的复制，而且还能够指导生物的发育以及细胞的日常活动呢？遗传信息包含什么样的指令？这些指令的物理组织又是怎样的，以至于单个细胞狭小的空间内能够储存发育和维持所需的大量信息，即使是最简单的生物？

其中一些问题的答案在 20 世纪 40 年代开始出现，通过对简单真菌的研究，这个时期研究者们发现，遗传信息主要由生产蛋白质的指令组成。蛋白质是行使大多数细胞功能的大分子：它们可以作为细胞结构的构造元件，也可以形成催化所有细胞化学反应的酶（讨论见第 3 章），还可以调节基因表达（讨论见第 7 章），使细胞能够运动（讨论见第 6 章）并能在彼此间进行通信（讨论见第 15 章）。细胞的特性与功能几乎全部由它们所能合成的蛋白质决定。再加上后来的认识，就很难想像到遗传信息还曾含有什么其他类型的指令。

20 世纪 40 年代获得的另一关键进展是证实了**脱氧核糖核苷酸（DNA）**是遗传信息可能的载体，但是对于遗传信息从一个细胞传递到另一个细胞所需的复制，以及 DNA 的结构如何指令蛋白质合成的机制问题仍然完全不清楚。1953 年，由于 James Watson 和 Francis Crick 发现了 DNA 的结构，这个难题意外地得以解决。第 1 章提到过，DNA 的结构立刻解决了该分子内信息怎样拷贝，或者复制这一难题。它也同时为发现 DNA 分子如何编码生产蛋白质的指令提供了线索。今天，DNA 是遗传物质对于生物学家来讲是如此基本的事实，以至于难以想像到这个发现填补了多么巨大的知识空白。

在生物学家发现 DNA 的结构以前，他们就认识到基因由**染色体**携带，在 20 世纪人们发现它们以线状结构存在于真核细胞的核内，当细胞开始分裂时，就变得清晰可见（图 4-1）。后来，由于得以进行生物化学分析，才发现染色体由 DNA 和蛋白质组成。现在我们知道 DNA 携带了细胞的遗传信息（图 4-2）。相比之下，染色体的蛋白质组分主要起包装以及控制冗长的 DNA 分子的作用，这样就可以将这些 DNA 固定在细胞内，并在需要时容易接触到它们。

本章我们首先描述 DNA 的结构。我们看看 DNA 的结构和化学特性是如何使其成为理想的基因原材料的，尽管它的化学很简单。地球上每个细胞的基因都由 DNA 组成，对 DNA 和

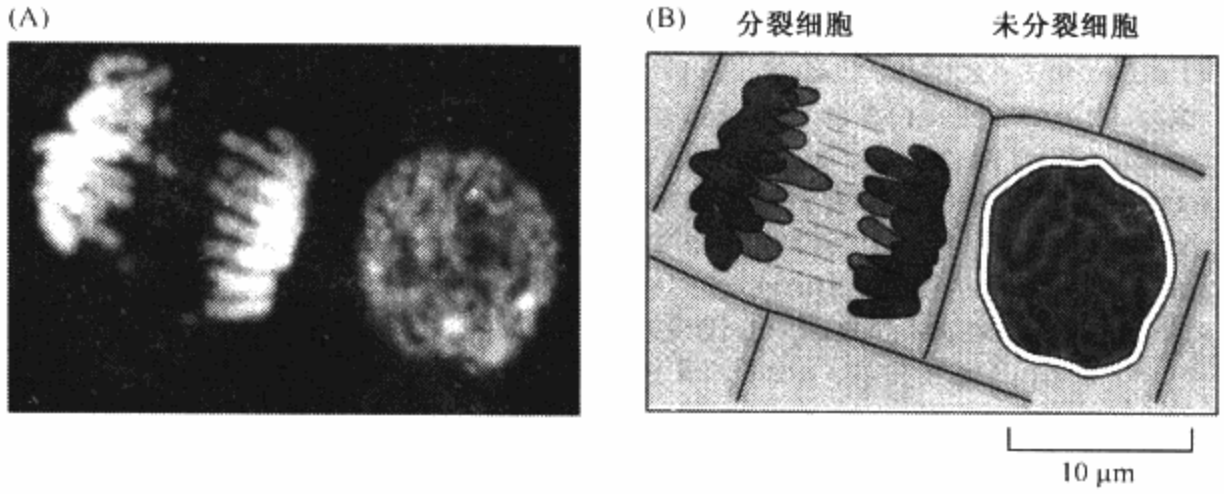


图 4-1　细胞内的染色体

（A）光学显微镜下拍摄到的两个相邻植物细胞。DNA 已用荧光染料（DAPI）染色，这种染料可以结合 DNA。DNA 位于染色体上，只有在准备进行细胞分裂时，成为致密的结构以后，染色体才能在光学显微镜下清楚地看到是不同的结构，如（A）图所示。右边是没有分裂的细胞，含有相同的染色体，但是在细胞生命周期的这一阶段，在光学显微镜下不能清晰地辨别它们；因为它们处于更加伸展的构象。（B）两个细胞及其染色体轮廓的简图[（A）经 Peter Shaw 许可]

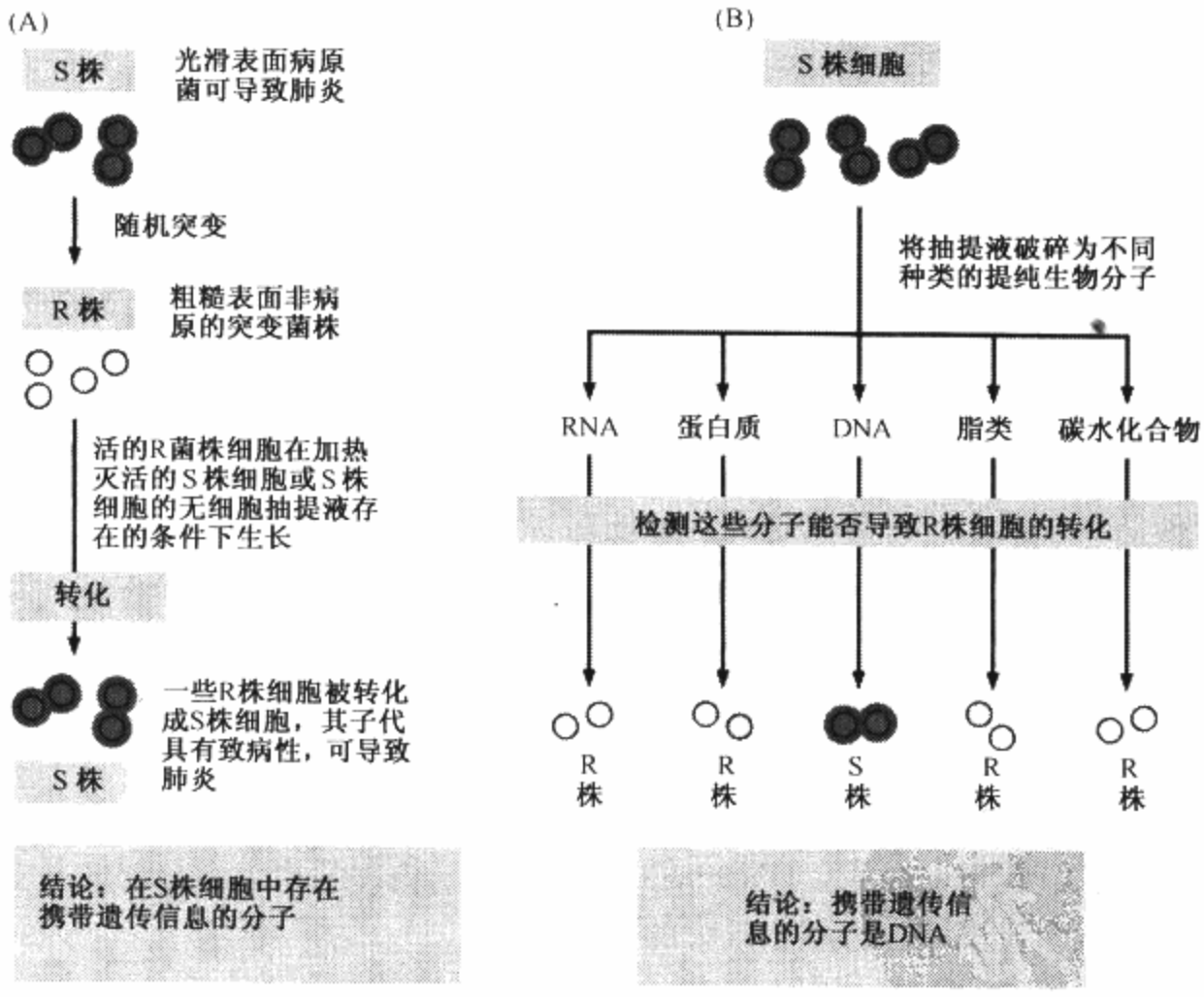


图 4-2　实验证明 DNA 是遗传物质

这些在 20 世纪 40 年代进行的实验表明，将纯化的 DNA 加入到细胞中，会改变它的特性，而且这种改变能忠实地遗传到下一代。肺炎球菌两种密切相关的菌株之间在显微镜下的表型以及病原性方面存在差别。一个菌株表面光滑（S），注射到老鼠体内后会导致死亡，另一个菌株表面粗糙（R）且不致命。（A）该实验表明 S 株内存在的物质能使 R 株改变（或转化）成 S 株，并且这种改变能遗传给下一代细菌。（B）该实验中，从 S 株获得的不同生物分子与 R 株培养以后，确定了该物质是 DNA

基因之间关系的深刻认识来自于在多种生物体内进行的实验研究。然后我们考察基因和其他重要的DNA 片段如何排列在染色体中冗长的 DNA 分子上。最后，我们讨论真核细胞如何将这些冗长的分子折叠成致密的染色体。这种包装必须以有序的方式进行，以便染色体在每次细胞分裂时能够正确复制并在两个子细胞间正确分配。这种组装也必须让染色体 DNA 发生损伤时修复它们的酶，以及指导染色体 DNA 许多基因表达的特殊蛋白质能够接触到它。

本章是 4 个讨论基础遗传机制（细胞维持、复制、表达以及偶尔改进它的 DNA 内携带的遗传信息的方式）章节中的第一章。下一章（第 5 章），我们讨论细胞精确地复制和修复 DNA 的机制；我们还描述 DNA 序列如何通过遗传重组过程进行重排。基因表达（细胞解译编码在 DNA 内的信息来指导蛋白质合成的过程）则是第 6 章讨论的主题。第 7 章我们描述细胞如何控制基因表达，以确保在细胞生命过程中，它的 DNA 编码的成千上万种蛋白质每一种都只在正确的时间和地点合成出来。在有关基础遗传机制的这 4 个章节后面，紧接着我们讲述了用于研究以上提到的和其他细胞内基本过程（讨论见第 8 章）的实验技术。

DNA 的结构和功能

20 世纪 40 年代，由于 DNA 的化学从表面上看极为简单，生物学家难以接受将 DNA 当作遗传物质。DNA 是一种仅由 4 种在化学上彼此相似的亚单位组成的长链多聚物。20 世纪 50 年代早期，首次用一种确定分子三维原子结构的技术，X 射线衍射分析，研究了 DNA（讨论见第 8 章）。早期的 X 射线衍射分析结果表明，DNA 由两股聚合物组成，它们彼此盘绕形成一个双螺旋。观察到 DNA 是双链的具有极其重要的意义，它为 Watson-Crick DNA 结构的发现提供了主要的线索。只有当这种结构模型提出以后，DNA 进行复制和编码信息的潜在能力才显现出来。这一部分我们考察 DNA 分子的结构并从总体上解释它是如何能够储存遗传信息的。

DNA 分子由两条互补的核苷酸链组成

4 种核苷酸亚单位构成一条多聚核苷酸长链，两条多聚核苷酸长链再组成一条 DNA 分子。其中每一条称为一条 **DNA 链**或者 **DNA 股**。核苷酸的碱基部分之间的**氢键**将这两条链维系在一起（图 4-3）。第 2 章（专题 2-6）中我们看到，核苷酸由一个结合有单个或更多磷酸基团的五碳糖和一个含氮碱基组成。对于 DNA 中的核苷酸，这个糖是连有一个磷酸基因的脱氧核糖（因而得名脱氧核糖核苷酸），碱基则可以是**腺嘌呤**（A）、**胞嘧啶**（C）、**鸟嘌呤**（G）或**胸腺嘧啶**（T）。核苷酸通过糖和磷酸之间以共价键连接成一条链，因此它们形成了一条交替出现的糖-磷酸-糖-磷酸骨架（图 4-3）。由于这 4 种亚单位中，仅仅碱基不同，因此 DNA 中的每一条多聚核苷酸链类似一条串有 4 种珠子（4 种碱基 A、C、G 和 T）的项链（骨架）。相同的这些符号（A、C、G、T）通常也用来代表 4 种不同的核苷酸，即连有糖和磷酸基因的碱基。

核苷酸亚单位线性排列的方式使 DNA 链具有了化学极性。如果我们把每个糖看成是一个一端有一个突出圆球（5′磷酸），另一端有一个凹陷的方块（图 4-3），那么由突出的圆球和凹陷连锁形成的完整核苷酸链，将使其所含全部亚单位按相同方向排列。而且，容易区别核苷酸链的 2 个末端，因为一端有一个凹陷（3′羟基），而另一端有一个圆球（5′磷酸）。把一端称为 3′端，另一端称为 5′端，用来表示 DNA 链中的这种极性。

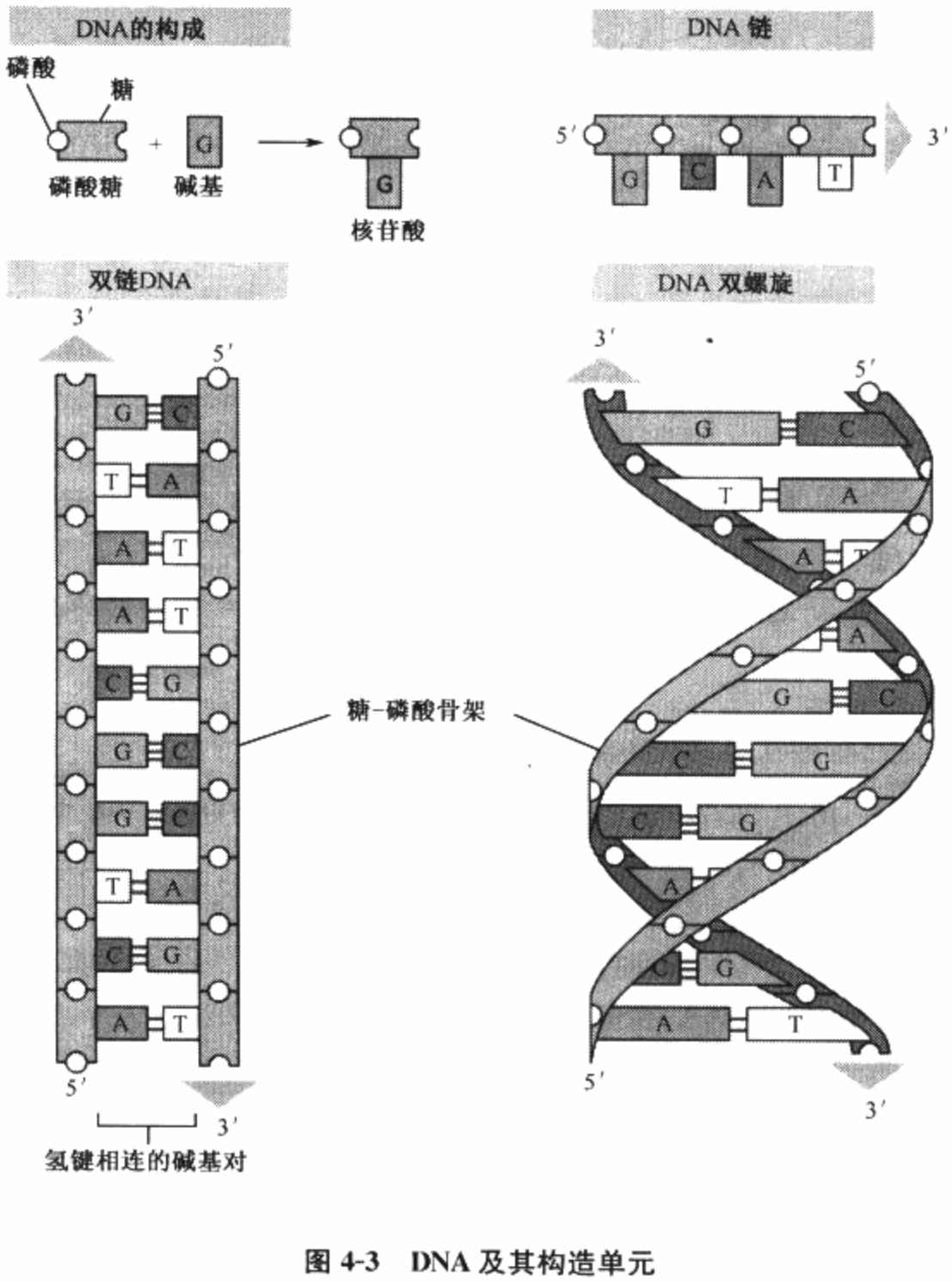


图 4-3　DNA 及其构造单元

DNA 由 4 种核苷酸组成，它们共价连接成一条多聚核苷酸链（一条 DNA 股），该链有一条糖-磷酸骨架，碱基（A、C、G 和 T）则从骨架上伸出。DNA 分子是由两条 DNA 链通过碱基对之间的氢键维系起来而形成。DNA 链末端的箭头表示分子内反向平行的两股 DNA 的极性。左下方的图中示出了笔直形式的 DNA 分子；实际上，它盘绕成一个双螺旋，如右下方的图中所示。详见图 4-5

DNA 两条多聚核苷酸链的化学和结构特征导致产生它的三维结构——**双螺旋**。由于这两条链通过不同链上碱基间的氢键维系在一起，所以所有的碱基都位于双螺旋的内侧，而糖-磷酸骨架位于外侧（图 4-3）。在每种情况下，一个体积较大的双环碱基（一个嘌呤，专题 2-6）与一个单环碱基（一个嘧啶）配对；A 通常与 T 配对，G 通常与 C 配对（图 4-4）。这种**互补碱基配**对使**碱基对**能够以能量上最有利的方式排列在双螺旋内部。在这种排列中，每一个碱基对都有相似的宽度，因而使糖-磷酸骨架沿 DNA 分子保持相同的距离。为了使碱基配对组装的效率最大化，两个糖-磷酸骨架彼此盘绕，形成一个双螺旋结构，每一圈完整螺旋含有 10 个碱基对（图 4-5）。

仅仅当双螺旋中的两条链反向平行时，也就是说，只有一条链的极性方向与另一条链的极性方向相反时每个碱基对中的成员才能恰好配对在一起（图 4-3 和图 4-4）。这些碱基配对

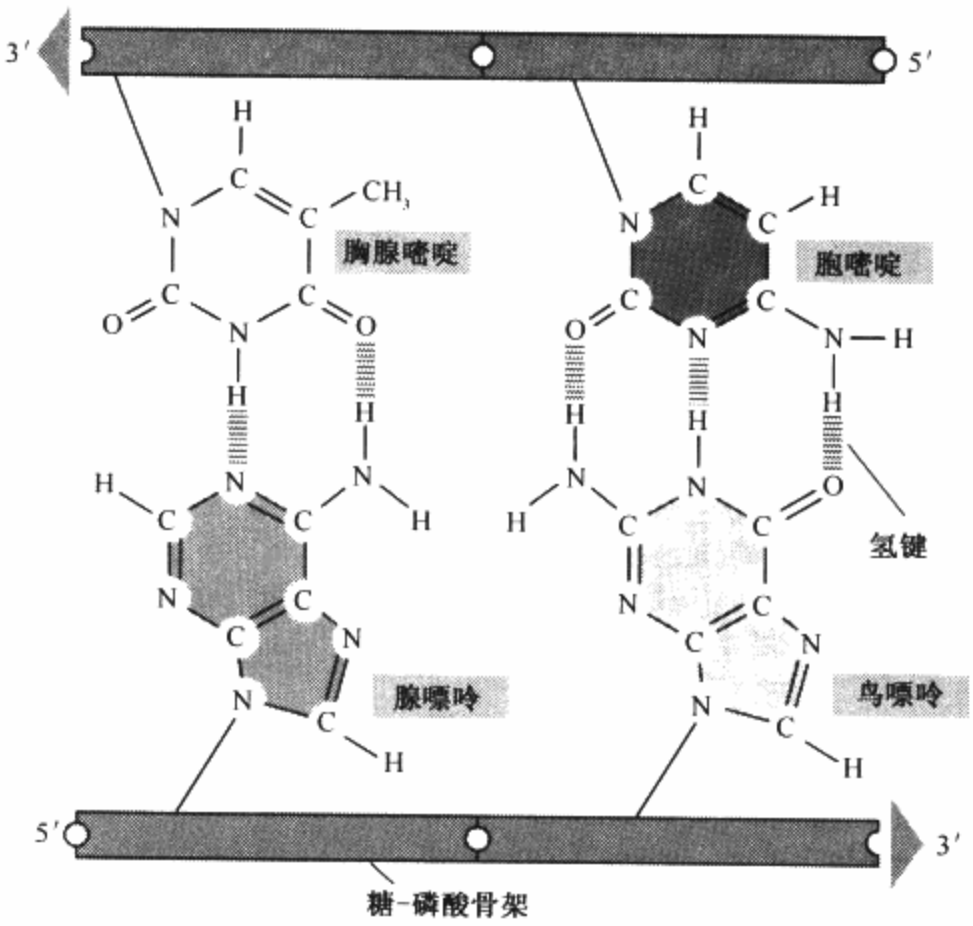


图 4-4　DNA 双螺旋中的互补碱基对

碱基的形状和化学结构只允许在 A 和 T 以及 G 和 C 之间有效地形成氢键，在这里能够形成氢键的原子可以彼此靠近却不会使双螺旋的形状发生扭曲（专题 2-3）。如图所示，A 和 T 之间形成 2 个氢键，而 G 和 C 之间形成 3 个氢键。含有碱基的两条多核苷酸链只有彼此反向平行时，这些碱基才能以这种方式进行配对

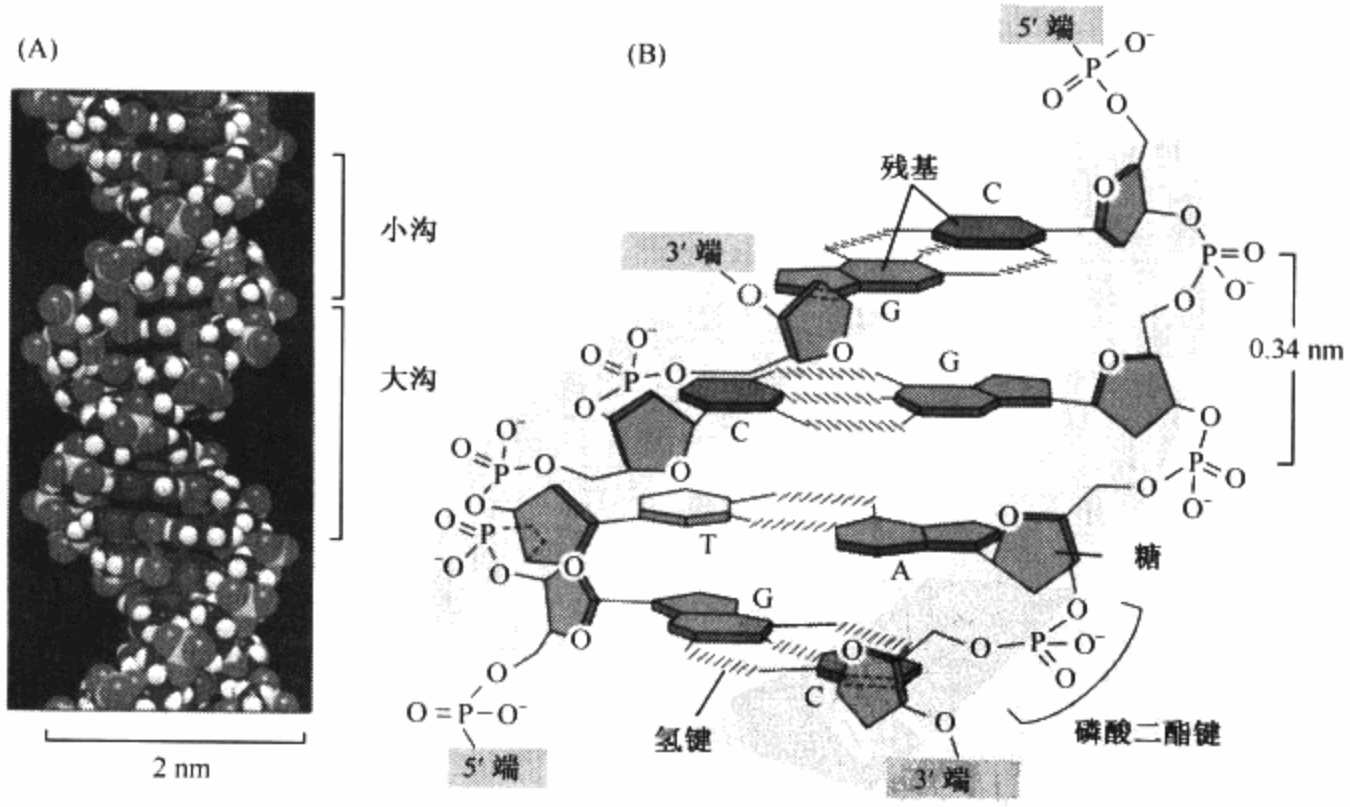


图 4-5　DNA 双螺旋

（A）1.5 圈 DNA 双螺旋的空间充满模型。每一圈 DNA 螺旋含有 10.4 个核苷酸对，相邻核苷酸对的中心间距是 3.4nm。两股 DNA 彼此间的盘绕在双螺旋中形成 2 个沟。如图所示，较宽的沟称为大沟，较窄的称为小沟。（B）从侧面看到的一小段双螺旋，显示了 4 个碱基对。核苷酸通过上一个糖的 3′-羟基（—OH）基团和下一个糖的 5′磷酸基团之间形成磷酸二酯键共价连接起来。因此每股多聚核苷酸有一个化学极性；也就是说它的两个末端从化学角度看是不同的。3′端携带一个连接在糖环上 3′位置的自由羟基，而 5′端携带一个连接在糖环上的 5′自由磷酸基团

要求产生的结果就是 DNA 分子的每一条链含有的核苷酸序列与另一条链上的核苷酸序列恰好**互补**。

DNA 的结构提供了一种遗传机制

携带生物信息的基因，为了传递给下一代，在细胞每次分裂形成两个子细胞时，都必须进行精确地拷贝。这些要求产生了两个重要的生物学问题：指令生物体的信息是怎样以化学形式携带的？又是怎样精确地拷贝的？DNA 双螺旋结构的发现是 20 世纪生物学研究中的一个里程碑，因为这个发现立即暗示了这两个问题的答案，从而在分子水平上解答了这个遗传难题。这一部分，我们简要地讨论这些问题的答案，在后面的章节我们还将进行更详细的分析。

DNA 通过沿两条链排列的核苷酸的顺序或序列编码遗传信息。每个碱基——A、C、T 或 G 都可以当作是编写出 DNA 化学结构内生物信息的四字母表中的一个字母。第 1 章我们看到，生物体之所以彼此不同，是因为它们各自的 DNA 分子拥有不同的核苷酸序列，因而能够携带不同的生物信息。但是，怎样用核苷酸字母表来构成信息？这些信息又具体说明了什么呢？

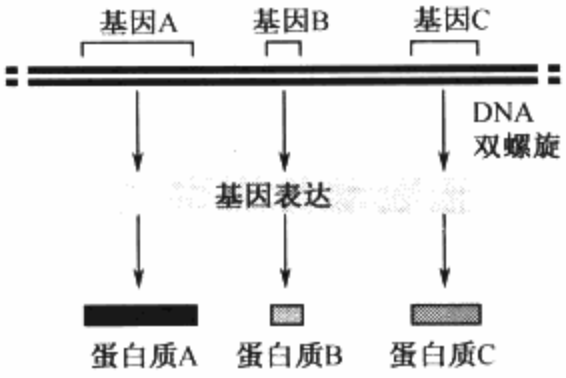


图 4-6　携带在 DNA 中的遗传信息与蛋白质之间的关系

正如前面所讨论的，在发现 DNA 结构以前，就已经很清楚基因包含着合成蛋白质的指令。因此在某种程度上，DNA 信息必须编码蛋白质（图 4-6）。鉴于蛋白质的化学特性，这种关系立即就使得前面提到的难题更加容易理解。第 3 章讨论过，蛋白质的生物学功能由其特性决定，而其特性由它的三维结构决定它的结构又由组成它的氨基酸线性序列决定。因此，基因内核苷酸的线性序列在某种程度上必须编写了蛋白质内氨基酸的线性序列。DNA 的四字母核苷酸字母表与蛋白质的 20 字母氨基酸字母表之间确切的对应关系（遗传密码）从 DNA 的结构来看并不明显，在 DNA 双螺旋结构发现以后，科学家们花费了 20 年的时间才弄清楚。第 6 章，在详尽阐明基因表达的过程中，我们也会详细描述这个密码，细胞通过基因表达的过程将基因内的核苷酸序列翻译成蛋白质的氨基酸序列。

生物体 DNA 完整的信息集合称为它的**基因组**，它携带了生物体能够合成的全部蛋白质编码信息。（基因组这个术语也用于描述携带这些信息的 DNA。）基因组内所含的信息量令人吃惊。例如，一个典型的人类细胞所含 DNA 长达 2m，如果用 4 字母核苷酸表编写出来，那么一段很小的人类基因核苷酸序列都将占一页纸内容的 1/4（图 4-7），而人类基因组的完整核苷酸序列如果写成与本书同样大小的书，数量将超过 1000 本。除了其他重要信息以外，基因组携带了编码大约 30 000 种不同蛋白质的指令信息。

细胞在每一次分裂时，为了将它的基因组传递到 2 个子细胞中，都必须对基因组进行拷贝。DNA 结构的发现同样也揭示出这种拷贝得以进行的原理：由于 DNA 的每一条链含有的核苷酸序列与另一条链的核苷酸序列恰好互补，所以每一条链都可以作为**模板**或模子，用来合成新的互补链。换句话说，如果我们将两条 DNA 链命名为 S 和 S′，那么 S 链可以作为合成新的 S′链的模板，而 S′链也可以作为合成新的 S 链的模板（图 4-8）。因此，通过这种完美而又简单的过程，即 S 链与 S′链分离，然后每一条分离的链作为生成新的互补配对链的模板，能够精确地拷贝 DNA 中的遗传信息。

CCCTGTGGAGCCACACCCTAGGGTTGGCCA
ATCTACTCCCAGGAGCAGGGAGGGCAGGAG
CCAGGGCTGGGCATAAAAGTCAGGGCAGAG
CCATCTATTGCTTACATTTGCTTCTGACAC
AACTGTGTTCACTAGCAACTCAAACAGACA
CCATGGTGCACCTGACTCCTGAGGAGAAGT
CTGCCGTTACTGCCCTGTGGGGCAAGGTGA
ACGTGGATGAAGTTGGTGGTGAGGCCCTGG
GCAGGTTGGTATCAAGGTTACAAGACAGGT
TTAAGGAGACCAATAGAAACTGGGCATGTG
GAGACAGAGAAGACTCTTGGGTTTCTGATA
GGCACTGACTCTCTCTGCCTATTGGTCTAT
TTTCCCACCCTTAGGCTGCTGGTGGTCTAC
CCTTGGACCCAGAGGTTCTTTGAGTCCTTT
GGGGATCTGTCCACTCCTGATGCTGTTATG
GGCAACCCTAAGGTGAAGGCTCATGGCAAG
AAAGTGCTCGGTGCCTTTAGTGATGGCCTG
GCTCACCTGGACAACCTCAAGGGCACCTTT
GCCACACTGAGTGAGCTGCACTGTGACAAG
CTGCACGTGGATCCTGAGAACTTCAGGGTG
AGTCTATGGGACCCTTGATGTTTTCTTTCC
CCTTCTTTTCTATGGTTAAGTTCATGTCAT
AGGAAGGGGAGAAGTAACAGGGTACAGTTT
AGAATGGGAAACAGACGAATGATTGCATCA
GTGTGGAAGTCTCAGGATCGTTTTAGTTTC
TTTTATTTGCTGTTCATAACAATTGTTTTC
TTTTGTTTAATTCTTGCTTTCTTTTTTTTT
CTTCTCCGCAATTTTTACTATTATACTTAA
TGCCTTAACATTGTGTATAACAAAAGGAAA
TATCTCTGAGATACATTAAGTAACTTAAAA
AAAAACTTTACACAGTCTGCCTAGTACATT
ACTATTTGGAATATATGTGTGCTTATTTGC
ATATTCATAATCTCCCTACTTTATTTTCTT
TTATTTTTAATTGATACATAATCATTATAC
ATATTTATGGGTTAAAGTGTAATGTTTTAA
TATGTGTACACATATTGACCAAATCAGGGT
AATTTTGCATTTGTAATTTTAAAAAATGCT
TTCTTCTTTTAATATACTTTTTTGTTTATC
TTATTTCTAATACTTTCCCTAATCTCTTTC
TTTCAGGGCAATAATGATACAATGTATCAT
GCCTCTTTGCACCATTCTAAAGAATAACAG
TGATAATTTCTGGGTTAAGGCAATAGCAAT
ATTTCTGCATATAAATATTTCTGCATATAA
ATTGTAACTGATGTAAGAGGTTTCATATTG
CTAATAGCAGCTACAATCCAGCTACCATTC
TGCTTTTATTTTATGGTTGGGATAAGGCTG
GATTATTCTGAGTCCAAGCTAGGCCCTTTT
GCTAATCATGTTCATACCTCTTATCTTCCT
CCCACAGCTCCTGGGCAACGTGCTGGTCTG
TGTGCTGGCCCATCACTTTGGCAAAGAATT
CACCCCACCAGTGCAGGCTGCCTATCAGAA
AGTGGTGGCTGGTGTGGCTAATGCCCTGGC
CCACAAGTATCACTAAGCTCGCTTTCTTGC
TGTCCAATTTCTATTAAAGGTTCCTTTGTT
CCCTAAGTCCAACTACTAAACTGGGGGATA
TTATGAAGGGCCTTGAGCATCTGGATTCTG
CCTAATAAAAAACATTTATTTTCATTGCAA
TGATGTATTTAAATTATTTCTGAATATTTT
ACTAAAAAGGGAATGTGGGAGGTCAGTGCA
TTTAAAACATAAAGAAATGATGAGCTGTTC
AAACCTTGGGAAAATACACTATATCTTAAA
CTCCATGAAAGAAGGTGAGGCTGCAACCAG
CTAATGCACATTGGCAACAGCCCCTGATGC
CTATGCCTTATTCATCCCTCAGAAAAGGAT
TCTTGTAGAGGCTTGATTTGCAGGTTAAAG
TTTTGCTATGCTGTATTTTACATTACTTAT
TGTTTTAGCTGTCCTCATGAATGTCTTTTC

图 4-7　人类 β-珠蛋白基因的核苷酸序列

该基因携带血红蛋白分子两种亚基其中一个的氨基酸序列编码信息，血红蛋白的作用是在血液中运输氧气。另一个不同的基因，α-珠蛋白基因，携带血红蛋白另一种亚基的编码信息（每个血红蛋白分子含有 4 个亚基，每种类型各两个）。图中只显示了含有 β-珠蛋白基因 DNA 双螺旋的两股中的一股；另一股拥有恰好互补的序列。根据惯例，书写核苷酸序列时是从它的 5′端到 3′端，阅读时则是从左至右，沿直线顺着该页连续向下，就好像它是普通的英文字母一样。用黄色标注的 DNA 序列表示编码 β-珠蛋白氨基酸序列基因的 3 个区域。第 6 章我们将看到细胞怎样将这三段序列连接起来合成一个完整长度的 β-珠蛋白

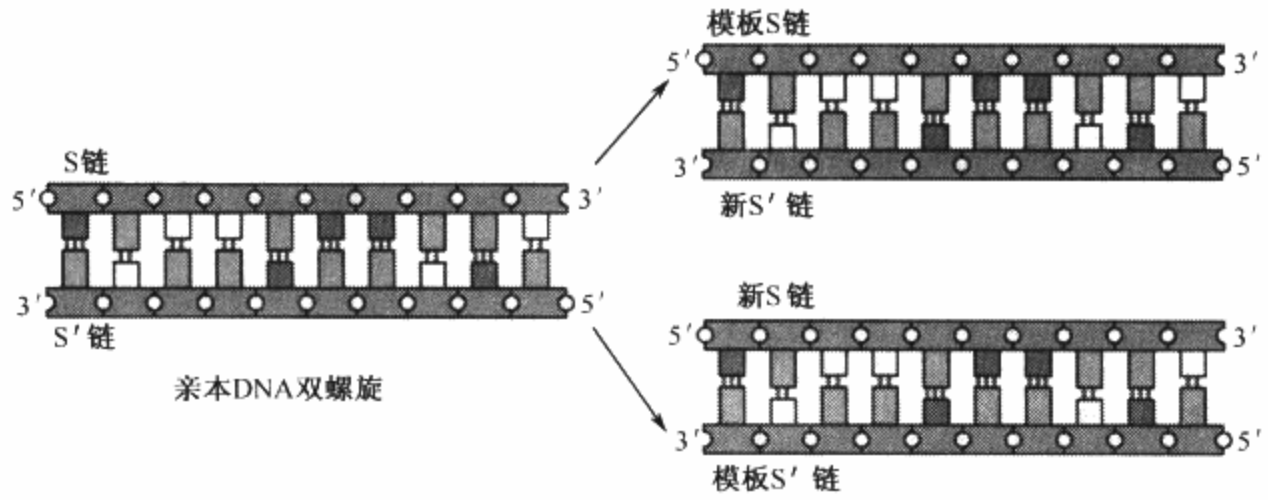


图 4-8　DNA 作为自身复制的模板

由于核苷酸 A 只能与 T 成功配对，G 与 C 配对，所以每股 DNA 能够指定其互补股上核苷酸序列。通过这种方式，双螺旋 DNA 能够进行精确地复制

DNA 分子的每一条链作为模板合成互补链的能力，使得细胞在将基因传递给它的后代时，能拷贝或者**复制**它的基因。下一章，我们描述细胞完成这项艰巨任务的精巧机制。

真核生物中，DNA 围在细胞核内

真核细胞内几乎所有 DNA 都隔离在约占细胞总体积 10%的细胞核内。两层同心脂双层膜形成的**核被膜**界定了该区室，而在细胞核和细胞质之间转运分子的巨大核孔间隔穿过了这两层膜。核被膜直接与内质网大面积的膜相连。它由两层中间纤维网络提供机械支持：一个称作**核纤层**，在核内形成薄薄的片层网状物，就位于核内膜下面；另一个围绕着核外膜，且没有太多的组织有序性（图 4-9）。

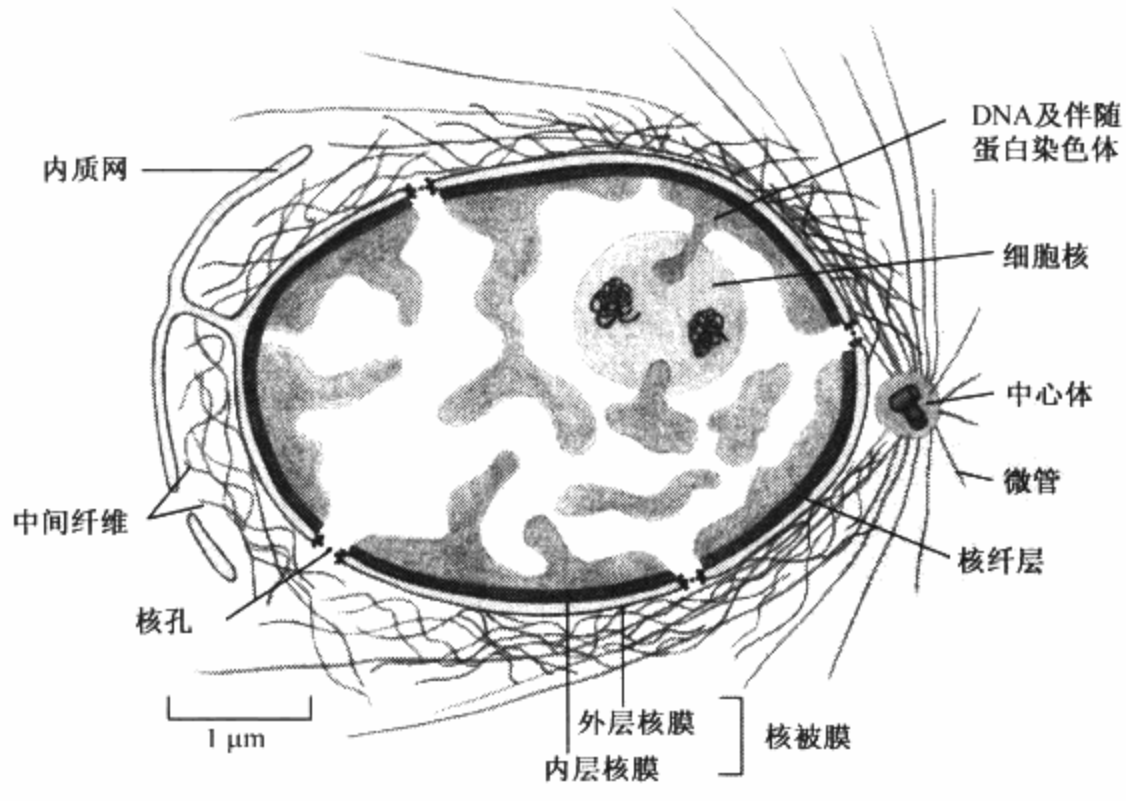


图 4-9　典型细胞核的横截面图

核被膜由两层膜组成，外层核膜与细胞内质网通连（图 12-9）。内质网内部空间（ER 内腔）用黄色表示；它与两层核膜之间的空间是连续的。内层核膜与外层核膜的脂双层在每一个核孔处相连。两个中间纤维（绿色表示）网络为核被膜提供机械支撑；细胞核内的纤维形成一种特殊的支撑结构称为核纤层

核膜使许多作用于 DNA 的蛋白质能够在细胞内需要它们的地方集中起来，而且我们在下章将看到，也使得细胞与细胞质分开，这个特点对于真核细胞发挥正常功能是至关重要的。区室化是生物学中一个重要的规则，细胞核就是一个例子；它帮助建立一种环境，由于存在高浓度的底物以及作用于它们的酶，所以大大促进了其中进行的生物化学反应。

小结

遗传信息携带在 DNA 核苷酸的线性序列中，每个 DNA 分子是一个双螺旋，由两条互补的核苷酸链组成，它们通过 G-C 和 A-T 碱基对之间的氢键维系在一起。遗传信息复制需要利用一条 DNA 链作为模板来形成一条互补链，生物体 DNA 中的遗传信息含有这个生物体所能合成的全部蛋白质的指令。在真核生物中，DNA 位于细胞核内。

染色体 DNA 及其在染色质纤维中的包装

DNA 最重要的功能就是携带基因，也就是指令所有组成一个生物体的蛋白质的信息，包括在什么时候，在什么类型的细胞内，以多大的量合成每一种蛋白质。真核生物的基因组分成多条染色体，在这一部分，我们可以看到基因通常是如何排列在每条染色体上。此外，我们还描述使染色体能够精确地复制并从上一代传递到下一代的特殊 DNA 序列。

但是要阐明 DNA 是如何包装的，还是会面临严重的困难。每个人类细胞所含 DNA 如果伸展开来，从头到尾有将近 2m 长；而含有 DNA 的人类细胞核的直径却仅仅约为 6μm，从几何学上讲，这相当于将 40km（24 英里）长的极细的线装进一个网球内！这种复杂的组装 DNA 的工作由特殊的蛋白质完成，它们结合并折叠 DNA，从而形成一系列盘绕和环作为更高水平的组织形式，并防止 DNA 成为难控制的缠结。令人惊奇的是，尽管 DNA 折叠得非常紧密，其压缩的方式却使它容易为细胞内许多酶所利用，从而复制、修复它，并利用它的基因合成蛋白质。

真核生物 DNA 包装成一套染色体

在真核生物内，位于细胞核内的 DNA 分配在一组不同的**染色体**上。例如，人类基因组，大约 3.2×10^9 个核苷酸分布在 24 条不同的染色体上。每条染色体都含有一条极其冗长的与蛋白质结合的线性 DNA 分子，这些蛋白质将这条线状的 DNA 折叠并包装成更加致密的结构。DNA 和蛋白质形成的复合物称为**染色质**（来源于希腊语 chroma，即“有颜色的”，因为它可以被染色）。除参与包装 DNA 的蛋白质以外，染色体还结合有许多基因表达、DNA 复制，以及 DNA 修复过程所需的其他蛋白质。

细菌的基因由一条通常为环状的 DNA 分子携带（图 1-30）。这条 DNA 结合有蛋白质，它们参与组装、压缩 DNA，但是与真核生物中行使相同功能的蛋白质不同。尽管这条 DNA 通常也称作细菌的染色体，但是它并不具有与真核生物染色体相同的结构，至于细菌 DNA 是如何包装的所知较少。而古细菌环状 DNA 是如何压缩的则所知更少。因此，我们有关染色体结构的讨论，几乎全部围绕真核生物染色体进行。

除生殖细胞以及少数高度特化的，不能繁殖并缺少 DNA（如红细胞）的细胞类型之外，每个人细胞都含有每条染色体的两份拷贝，其中一份来自母亲，而另一份来自父亲。一对母源和父源染色体称为**同源染色体（同源物）**。仅男性有的一对非同源染色体是性染色体，其中的 **Y 染色体**来源于父亲，而 **X 染色体**来源于母亲。因此，每个人细胞总共含有 46 条染色体，

其中 22 对染色体在男性和女性中相同，再加上 2 条所谓的性染色体（在男性中为 X 和 Y，而在女性中为 2 条 X）。将每条染色体染成不同颜色后可以利用 DNA 杂交（详细描述见第 8 章）区分这些人类染色体（图 4-10）。染色体染色通常在细胞周期的特定阶段进行，在这时染色体高度凝聚，且容易观察到（有丝分裂，见下面）。

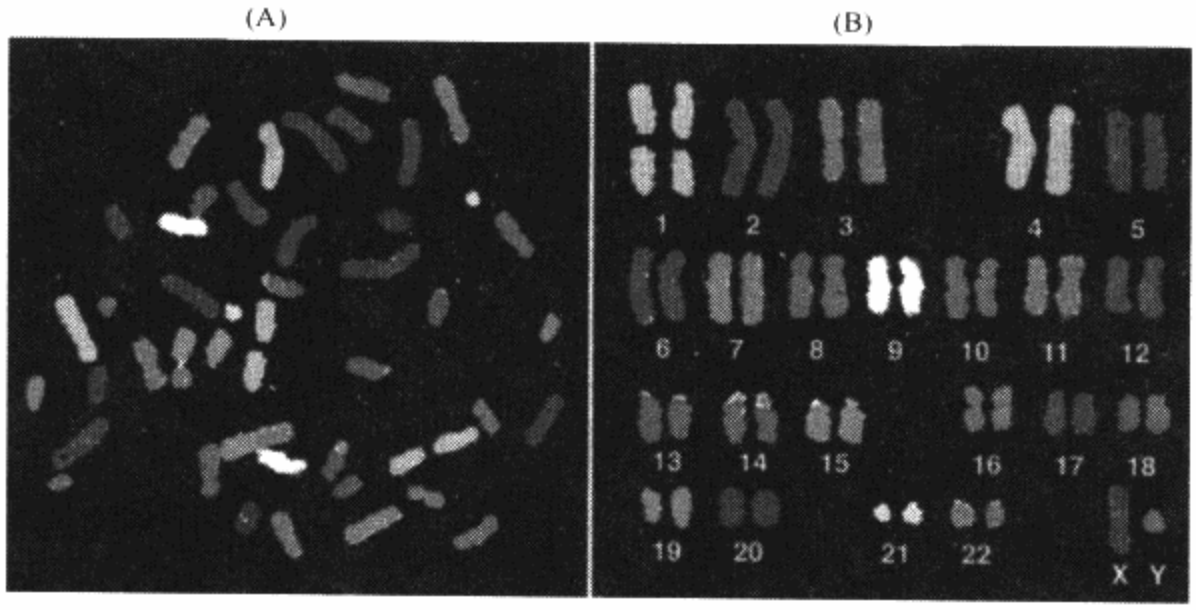


图 4-10　人类染色体

这些染色体来自男性，是从正在进行核分裂（有丝分裂）的细胞中分离得到，因此高度凝聚。每条染色体都染上不同颜色以便在光学显微镜下能清楚地加以分辨。将这些染色体和许多已偶联上不同荧光染料组合的人类 DNA 分子混合，就可以将它们染色。例如，来自 1 号染色体的 DNA 分子用一种特殊的染料组合标记，来自 2 号染色体的 DNA 分子则用另一种染料组合标记等。由于标记的 DNA 只能与所在的染色体（讨论见第 8 章）形成碱基配对，或者杂交，所以可对每条染色体进行不同的标记。对于这样的实验，要对染色体进行处理，将双链 DNA 解开成单链，以便能够与标记的单链 DNA 形成碱基配对，而同时仍然保持染色体结构完整性。（A）最初从裂解细胞中分离出来时观察到的染色体。（B）人为按标号顺序排列的同一套染色体。全套染色体的这种排列称作核型（摘自 E. Schrock et al., *Science* 273：494～497，1996. © AAAS）

另一个区分各条染色体的传统方法，是用能够产生沿每条有丝分裂染色体排列的显著可靠带型的染料将它们染色（图 4-11）。这些带型的结构基础还不太清楚，我们在本章最后再讨论这个问题。然而，每一种染色体上产生的带型都是是独一无二的，因此可以对各条染色进行鉴别并标号。

展示 46 条有丝分裂人类染色体的图称为人类的**核型**，如果其中一部分染色体发生丢失，或者彼此间发生交换，那么通过观察带型的改变或者染色体染色模式的改变，就可以发现这些变化（图 4-12），细胞遗传学家利用这些改变来发现染色体畸形，它们与遗传缺陷或者因体细胞内染色体重排产生某些癌症类型有关。

染色体含有长串的基因

染色体最重要的功能就是携带基因——遗传的功能单位。基因通常定义为含有合成一种特定蛋白质（或者许多紧密相关的蛋白质）指令的一段 DNA。尽管这个定义适用于大多数基因，但是占几个百分比的基因产生 RNA 分子（而不是蛋白质）作为它们终产物。这些 RNA 分子与蛋白质相似，在细胞内行使一组不同的结构和催化功能，我们将在后面的章节详细地进行讨论。

可能有人会想到，生物体的复杂性与其基因组内基因的数目存在相关性（表 1-1）。例如，生物体全部基因数量在简单细菌的 500 和人类的大约 30 000 之间变动，细菌和有些单细胞真

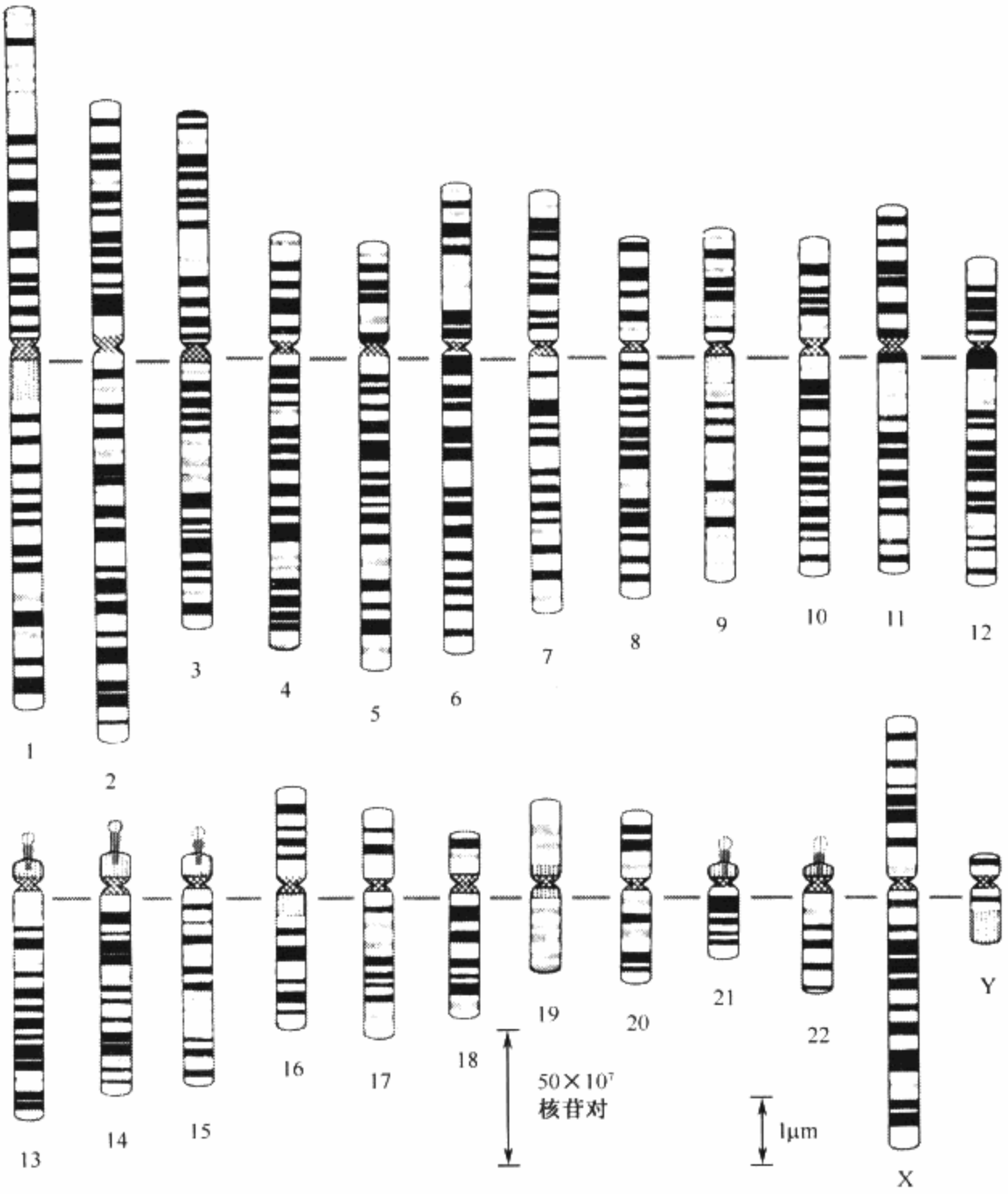


图 4-11　人类染色体带型

按照大致的大小顺序，将染色体从 1 到 23 编号，一个典型的人类体细胞（非生殖系）中，这些染色体每一条都有两份拷贝，再加上两条性染色体——女性有两条 X 染色体，男性有一条 X 染色体和一条 Y 染色体。用于作图的染色体在有丝分裂早期染色，此时染色体处于不完全压缩状态。水平的绿线代表着丝粒的位置（图 4-42），它就像有丝分裂染色体上的一个缢痕；13、14、15、21 和 22 号染色体上的球形突出物表示编码大核糖体 RNA 的基因所在位置（讨论见第 6 章）。用吉姆萨染液将染色体染色后获得这些带型，它们可以在光学显微镜下观察到（摘自 U. Franke, *Cytogenet. Cell Genet.*, 31：24～32，1981）

核生物具有特别紧凑的基因组；它们基因组完整核苷酸序列表明组成它们染色体的 DNA 分子全部都是紧密包装起来的成串基因（图 4-13 和图 1-30）。然而，许多真核生物（包括人类）的染色体除了含有基因以外，还有大量冗余的散在 DNA，它看起来并没有携带重要的信息。有时候，将这种 DNA 称作垃圾 DNA 以表示它对细胞的作用还没有被证明，其特定核苷酸序列可能并不重要；但是这种 DNA 本身，作为间隔材料，对于物种长期进化以及基因的正确表达是关键的，这些问题将在第 7 章详细论述。

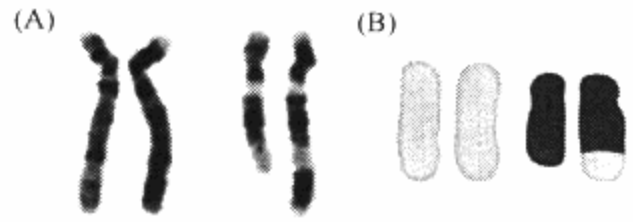


图 4-12　畸变的人类染色体

（A）这两对染色体已用吉姆萨染液染色（图 4-11），来自患有共济失调的患者，这种疾病的特征是运动技巧的逐渐退化。患者有一对正常的 4 号染色体（左边的一对），但一对 12 号染色体中一条正常，一条畸变，这可以从它的长度更长一些看出来（右边的一对）。根据其带型，可以推断 12 号畸变的染色体所含额外的物质是 4 号染色体的一小段，它通过一种称作染色体易位的异常重组事件附着到 12 号染色体上。（B）相同的两对染色体，蓝色的是 4 号染色体 DNA，紫色的是 12 号染色体 DNA。这两种技术产生的有关 12 号畸变染色体实质的结论是相同的，但染色体染色提供了更高的分辨率，并能清楚地鉴定出即便很小一段发生易位的染色体。然而吉姆萨染色更容易操作（摘自 E. Schrock et al., *Science* 273: 494～497, 1996, © AAAS）

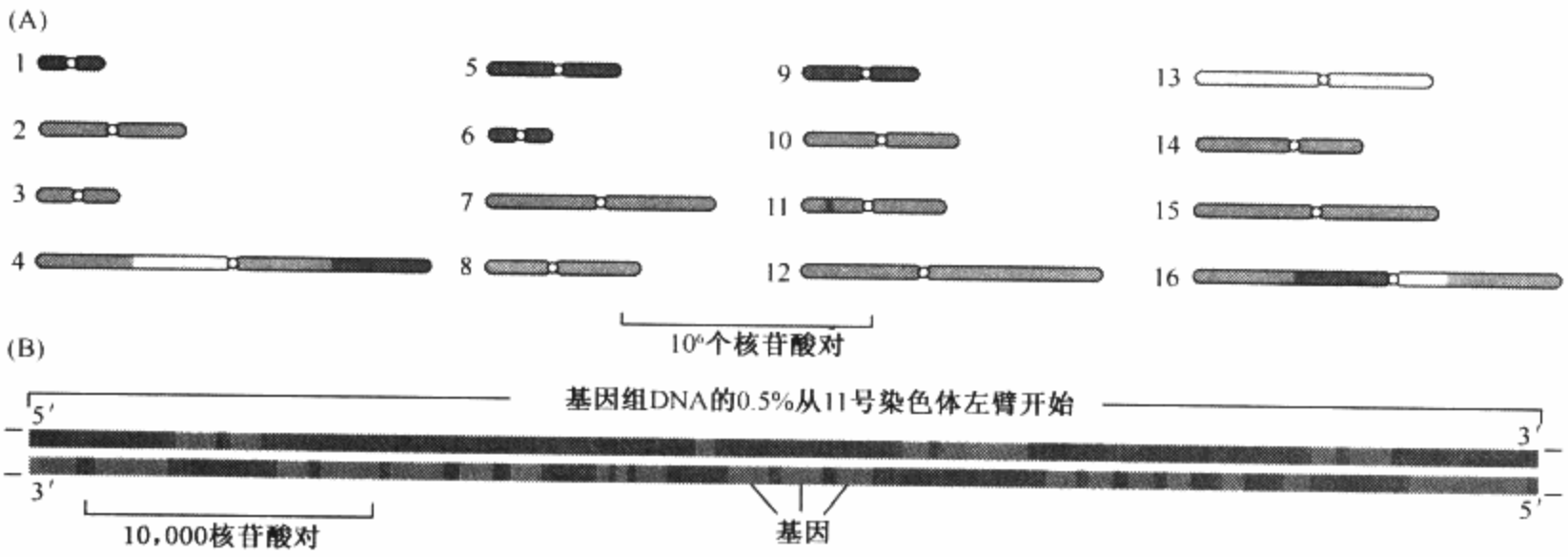


图 4-13　酿酒酵母（芽殖酵母）的基因组

（A）该基因组分配在 16 条染色体上。其完整的核苷酸序列的确定，是在许多不同地方工作的科学家合作努力的结果，如图所示（灰色，加拿大；橙色，欧盟；黄色，英国；蓝色，日本；淡绿色，St Louis，密苏里；深绿，斯坦福，加利福尼亚）。每条染色体上的缢痕代表它的着丝粒所在位置（图 4-42）。（B）将（A）图中用红色标示的 11 号染色体一小段区域放大，以显示该物种基因的高密度特征。如橙色所示，有些基因从下面的带转录（图 1-15），而另外一些从上面的带转录。完整的基因组长 12、147、813 个核苷酸对，含有大约 6000 个基因

总的来讲，生物体越复杂，它的基因组就越大，但是由于冗余 DNA 的数量不同，这种相关性并不系统（图 1-38）。例如，人类基因组比酿酒酵母基因组大 200 倍，却比有些植物和两栖动物基因组小 30 倍，比阿米巴的基因组小 200 倍。而且，由于冗余 DNA 数量的差别，即使它们基因数目大致相同，类似生物（如多骨鱼）的基因组相差也可达数百倍。看起来很明显，高等真核生物细胞携带大量的冗余 DNA 并没有什么不利，无论它们的作用是什么。

不同真核生物之间基因组在染色体上的分配也不相同。例如，与人类的 46 条染色体相比，来自一种小鹿的体细胞内只有 6 条染色体，而来自于鲤鱼的体细胞则含有 100 多条。对于起源上很相近的物种，即便基因组大小差不多，染色体数目和大小却相差很大（图 4-14）。因此染色体数目、物种复杂性和总基因组大小之间的关系并不那么简单。相反，由于选择压力的作用，现今物种各自都经历了独特的表面上看似随机的遗传事件进化历史后，才形成了它们现有的基因组和染色体。

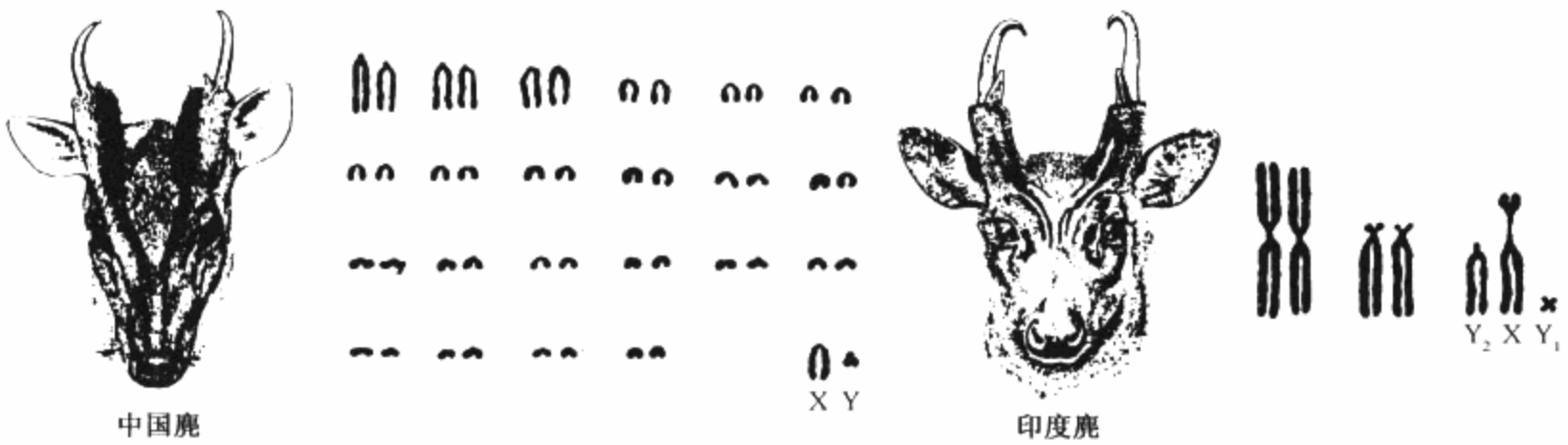


图 4-14　鹿的两个紧密相关的种具有非常不同的染色体数目

在印度鹿的进化过程中，起初分开的染色体融和到一起，但对这种动物没有很大的影响。这两个种的基因数目大致相同（摘自 M. W. Strickberger, Evolution, 3rd edition, 2000, Sudbury, MA: Johnes & Bartlett Publishers）

人类基因组的核苷酸序列揭示了基因在人体内是如何排列的

当 1999 年完成了人类的 22 号染色体，最小的人类染色体之一（图 4-11），测序的时候，首次可能看到基因在脊椎动物整条染色体上是如何排列的（表 4-1 和图 4-15）。2001 年发表了人类全部基因组的“第一张基因草图”后，人类全部染色体的遗传全貌立刻引起密切的关注。单单是人类基因组计划提供的信息量在生物学中就已史无前例了（图 4-16 和表 4-1）；人类基因组比任何其他到目前为止已完成测序的基因组大 25 倍，比已测序的基因组总和大 8 倍。人类基因组计划在最多的时候每天以 1000 个核苷酸/s 的速度得到初始核苷酸序列。彻底分析完这些信息还需要几十年，但是它将持续促进许多新实验方法的产生，而且它已经影响了本书中所有章节的内容。

表 4-1　人类 22 号染色体和全部人类基因组的详细统计

	22 号染色体	人类基因组
DNA 长度	48×10^6 核苷酸对	3.2×10^9
基因数量	700 左右	30 000 左右
编码蛋白的最小基因	1000 核苷酸对	未分析
最大基因	583 000 核苷酸对	2.4×10^6 核苷酸对
基因平均大小	19 000 核苷酸对	27 000 核苷酸对
一个基因外显子的最小数目	1	1
一个基因外显子的最大数目	54	178
一个基因外显子的平均数目	5.4	8.8
最小外显子大小	8 核苷酸对	未分析
最大外显子大小	7000 核苷酸对	17 106 核苷酸对
平均外显子大小	266 核苷酸对	145 核苷酸对
假基因数目	多于 134	未分析
位于外显子中的 DNA 所占百分数（编码蛋白质的序列）	3%	1.5%
位于高拷贝重复元件中的 DNA 所占百分数	42%	50%左右
占人类全基因组的百分数	1.5%	100%

33.8×10^6 核苷酸序列已知染色体的其余部分主要包含不编码蛋白质或 RNA 的很短的重复序列。

假基因是一段紧密装配的功能基因 DNA 的核苷酸序列，但含有大量缺失突变，阻碍了其正确表达。大多数假基因来源于有害突破积累在一个拷贝的功能基因复制。

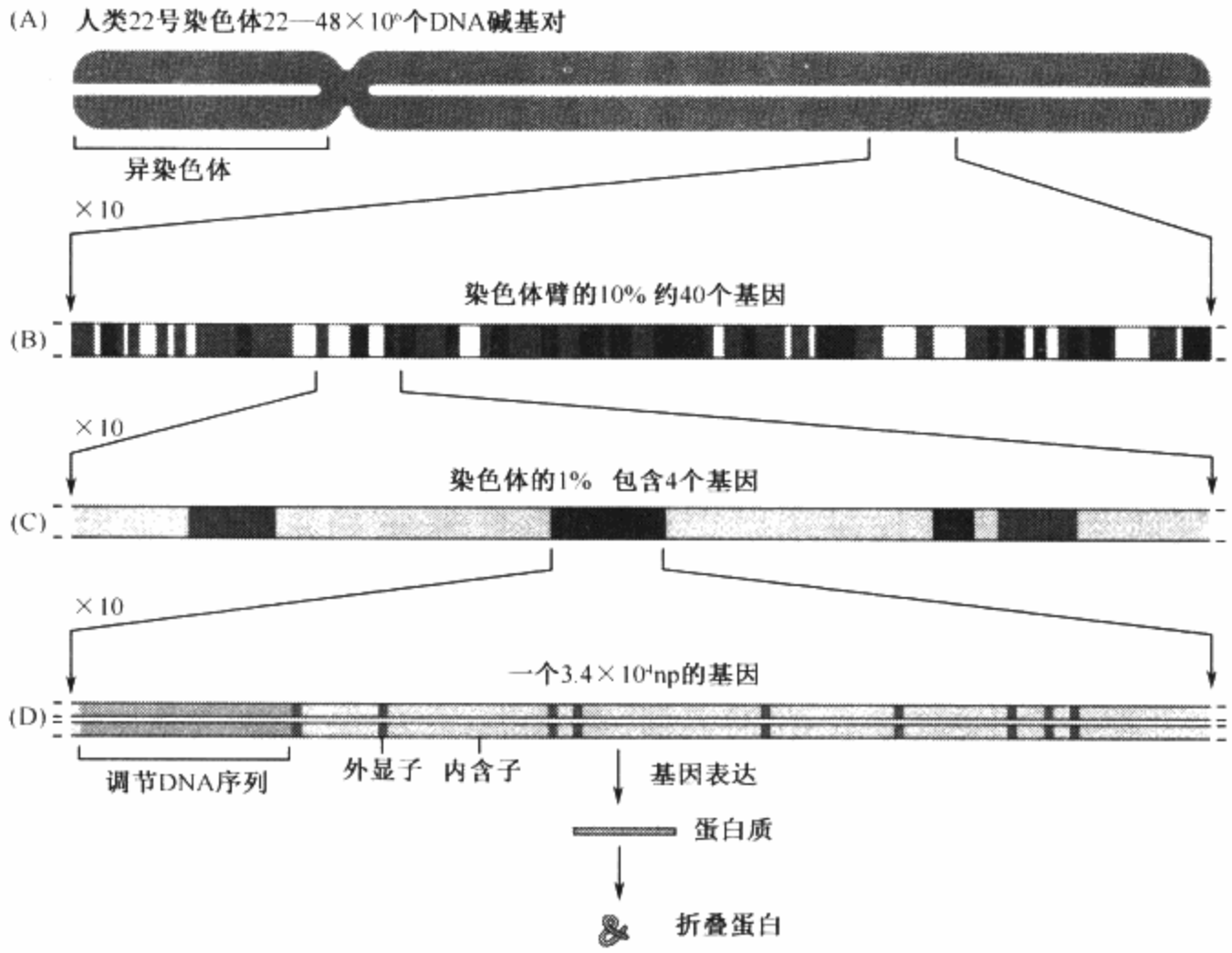


图 4-15　人类染色体上基因的组织

（A）22 号染色体，最小的人类染色体之一，含有 48×10^6 个核苷酸对，占整个人类基因组将近 1.5%，22 号染色体左臂大部分由短 DNA 重复序列组成，它们包装成特别紧密的染色质形式（异染色质），这在本章后面讨论。（B）一部分 22 号染色体的 10 倍放大图，显示了大约有 40 个基因。深褐色表示的是已知的基因，浅褐色表示的是估计的基因。（C）是（B）图的一部分放大后显示了几个基因的完整长度。（D）再放大 10 倍后显示了一个典型基因的内含子—外显子排列。每个外显子（红色）编码了蛋白质的一部分，而内含子 DNA 序列（灰色）相对不重要。整个人类基因组（3.2×10^9 个核苷酸对）分布在 22 条常染色体和 2 条性染色体上（图 4-10 和图 4-11）。人类基因组序列这个术语是指全部 24 条染色体中的 DNA 核苷酸序列。人类体细胞作为二倍体，因此含有 DNA 数量大致是这个数字的 2 倍。人类彼此之间平均每 1000 个核苷酸有 1 个不同，许多不同的人种为基因组测序计划提供了 DNA。因此发表的人类基因组序列是许多序列的综合（摘自 International Human Genome Sequencing Consortium，*Nature* 409：860～921，2001）

尽管可以从许多方面来分析人类基因组，这里我们仅仅就基因在人类染色体上的排列做一些概括。人类基因组第一个引人注意的地方就是编码蛋白质以及起结构和催化作用的 RNA 的序列如此少（仅占几个百分点）（图 4-17）。剩余的染色体 DNA 大部分都是一些短小且能转移的 DNA 片段，它们在进化过程中逐渐将自身插入到染色体中，我们将在后面的章节中详细讨论这些**转座元件**。

人类基因组另一个显著的特征就是平均每个基因大小多达 27 000 个核苷酸对。正如上面所讨论的，一个典型的基因在其线性核苷酸序列中携带了蛋白质氨基酸线性序列的编码信息。编码一个平均大小的蛋白质（人类大约是 430 个氨基酸）只需要大约 1300 个核苷酸对。基因中剩余的 DNA 大部分由长片段的非编码 DNA 序列组成，它们将编码蛋白质的 DNA 短片段

间隔开。这些编码序列称为外显子，而间隔（非编码）序列称为内含子（图 4-15 和表 4-1）。

由此可见，大多数人类基因都由长串交替排列的内含子和外显子组成，其中基因的大部分序列是内含子。相比之下，具有紧凑基因组生物的大多数基因缺少内含子。这是为什么它们的基因要小得多（大约是人类的 1/20），以及它们染色体中的编码 DNA 比例大得多的原因。除了内含子和外显子以外，每个基因还与调节 DNA 序列相连，它们负责确保基因以适当的水平，在适当的时间和正确的细胞类型中表达。在人类当中，一个典型基因的调控序列，长达数万核苷酸对。可以想到那些具有紧凑基因组的生物体内这些调控序列会更加压缩。

最后，人类基因组的核苷酸序列表明，那些关键的信息似乎处于惊人的混乱状态，正如一个评论家这样描述我们的基因组，“在某些方面它可能像你的车库、起居室、冰箱、生活：高度个性化，却又凌乱不堪；缺少组织；大多数像是积攒多年的杂物（非专业人士称其为‘垃圾’）；实际上不曾丢掉任何东西；少数明显有价值的元件不加区别且明显随便地分散在各处。”

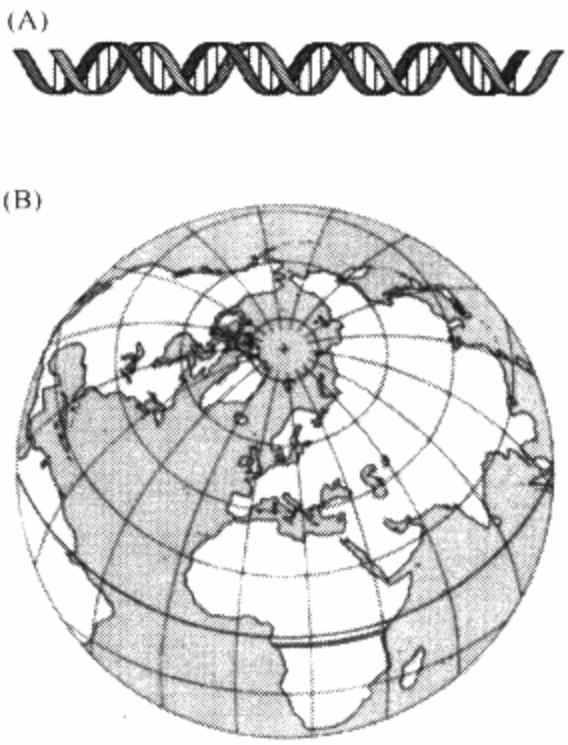


图 4-16　人类基因组大小

如果每个核苷酸对像（A）中这样画成 1mm 长，那么人类基因组将长达 3200km（将近 2000 英里），这足以横跨非洲中部，也就是我们人类起源的地方［（B）中用红色表示］。按这个比例平均来看，每 300m 会有一个蛋白质编码基因。一个中等大小的基因长达 30m，但该基因中的编码序列加起来不过 1m 多一点点

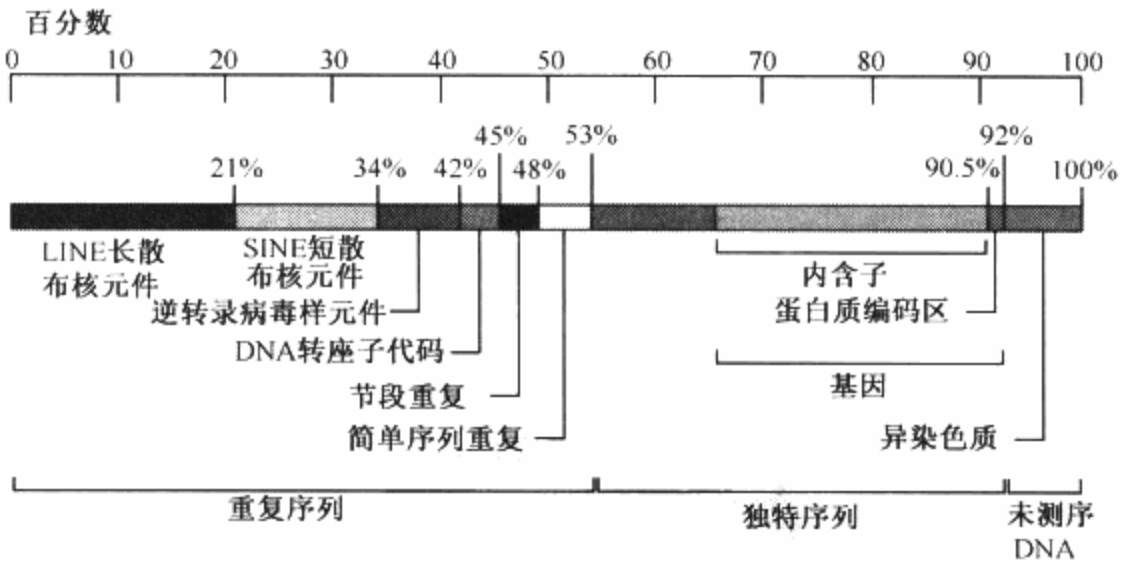


图 4-17　图解人类基因组核苷酸序列内容

长散布核元件、短散布核元件、逆转录病毒样元件、DNA 转座子都是可移动的遗传元件，它们在我们基因组中通过自身进行复制并将新的拷贝插入到不同位点来增殖。可移动遗传元件在第 5 章讨论。简单序列重复是在短核苷酸序列（不到 14 个核苷酸对），它们经过多次重复，形成大片段。节段重复是基因组中两个或更多位点都存在的大段基因组区域（1000～200 000 个核苷酸对）。独特序列中超过一半都由基因组成，其余的可能是调节 DNA。异染色质，一种基本不含基因的特殊染色质类型（本章后面讨论），其中存在的大部分 DNA 还没有测序（摘自 Unveiling the Human Genome, Supplement to the Wellcome Trust Newsletter. London: Wellcome Trust, February 2001）

相关生物DNA之间的比较揭示DNA序列中存在保守区域和非保守的区域

解释人类染色体核苷酸序列的一个主要困难就是大部分序列可能并不重要。而且基因组的编码区（外显子）发现通常都处于短片段（平均大小约145个核苷酸对）中，它们漂浮在无关紧要的核苷酸序列组成的DNA海洋中。这种排列使得鉴定一段DNA中所有的外显子相当困难，而要确定每个基因从哪里开始，又到哪里结束，以及它跨越多少个外显子，就难上加难了。精确的基因定位要求使用的方法能够从人类基因组固有的低信噪比背景中提取出有用信息，我们在将第8章中介绍其中的一些。这里我们讨论一种最普通的方法，它不仅可以鉴定出编码序列，而且能够确定其他重要的DNA序列。观察发现具有功能的序列在进化中保守，而没有任何功能的序列则能自由地进行随机突变，该方法正是基于这一点。因此其策略就是比较人类基因组序列和相关的某种生物，如小鼠的对应区域的序列。人类和小鼠据推

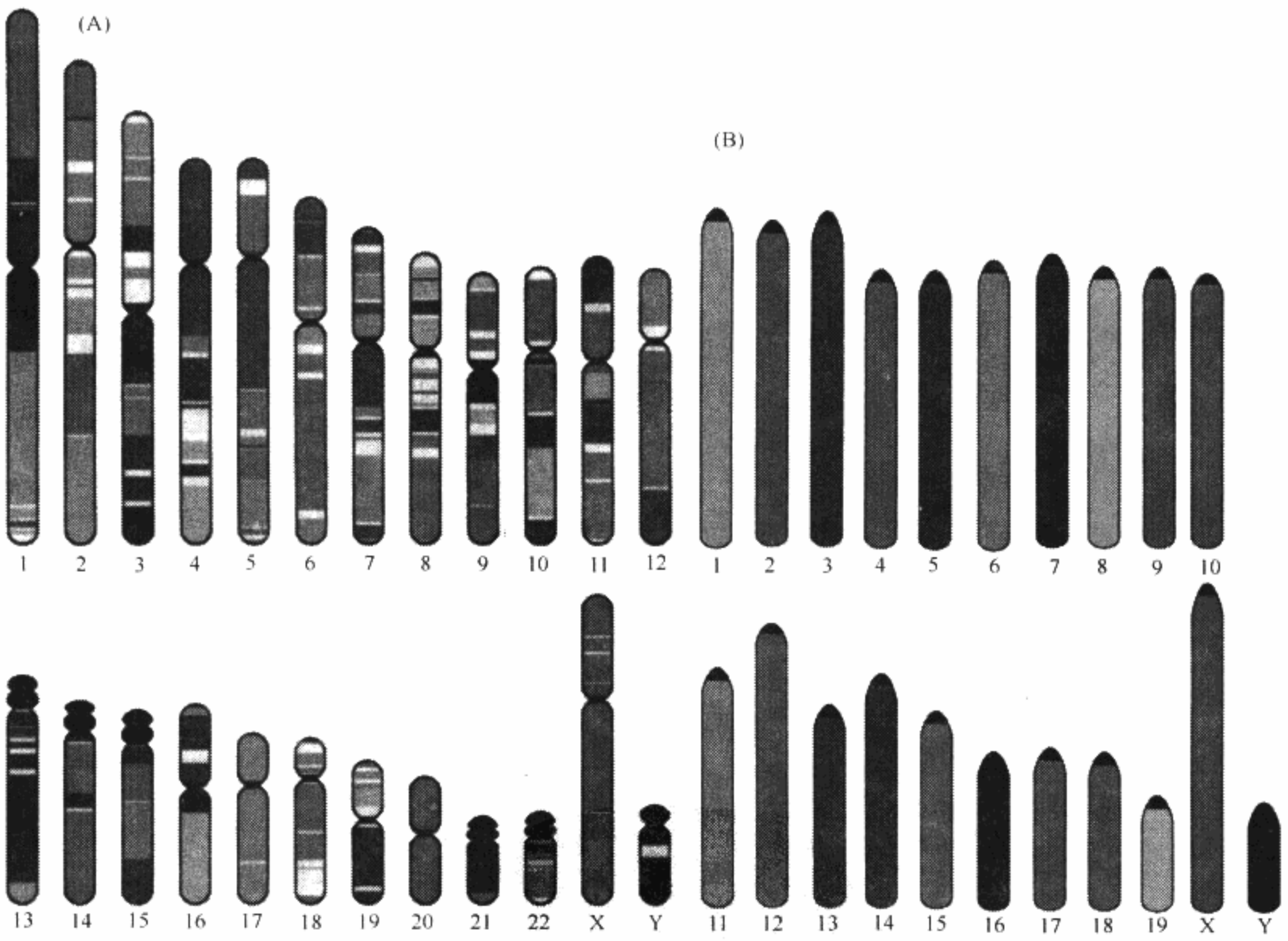


图 4-18　人类和小鼠基因组之间的保守同线性

小鼠染色体不同区域[(B)中用每只小鼠的颜色表示]表现出与所示的人类基因组（A）区域有保守的同线性（基因顺序）。例如，1号人类染色体（橙色）上部分中的基因与小鼠4号染色体中的一部分以相同顺序存在。黑色表示人类染色体中主要由短重复序列组成的区域。小鼠的着丝粒[(B)中用黑色表示]位于染色体末端；每条小鼠染色体上还没有已知基因位于着丝粒外面。对于大部分人类染色体而言，用缢痕表示的人类着丝粒占据染色体上靠中间的位置（图4-11）

（摘自 International Human Genome Sequencing Consortium，*Nature* 409：860～921，2001）

测大约在 10^8 年前就已从共同的哺乳动物祖先趋异，这么长的时间足以让它们基因中大多数核苷酸因随机突变事件而发生改变。结果，在两个基因组中仅有的仍然保持极其相似的区域，都是那样一些序列，一旦发生突变，将破坏基因的功能，使携带它们的动物在进化中处于不利地位，因而导致自然选择将它们从群体中淘汰。像这样非常相似的序列称为**保守区**。一般而言，保守区域代表功能上重要的外显子和调节序列。相反，**非保守区**代表通常对于功能并不关键 DNA 序列。由于通过这种方式揭示出了漫长自然“实验”的结果，所以 DNA 序列比较研究成为基因组学中最引人关注的领域。

这种比较研究不仅已经揭示了人类和小鼠拥有大部分的相同基因，而且小鼠和人类基因组大部分区段中，这些基因以相同的顺序排列，这个特征称为**保守同线性**（图 4-18）。染色体染色的方法也能揭示保守同线性，通过将我们人类染色体与其他哺乳动物染色体进行比较，这种技术已经用来再现我们人类染色体的进化历史（图 4-19）。

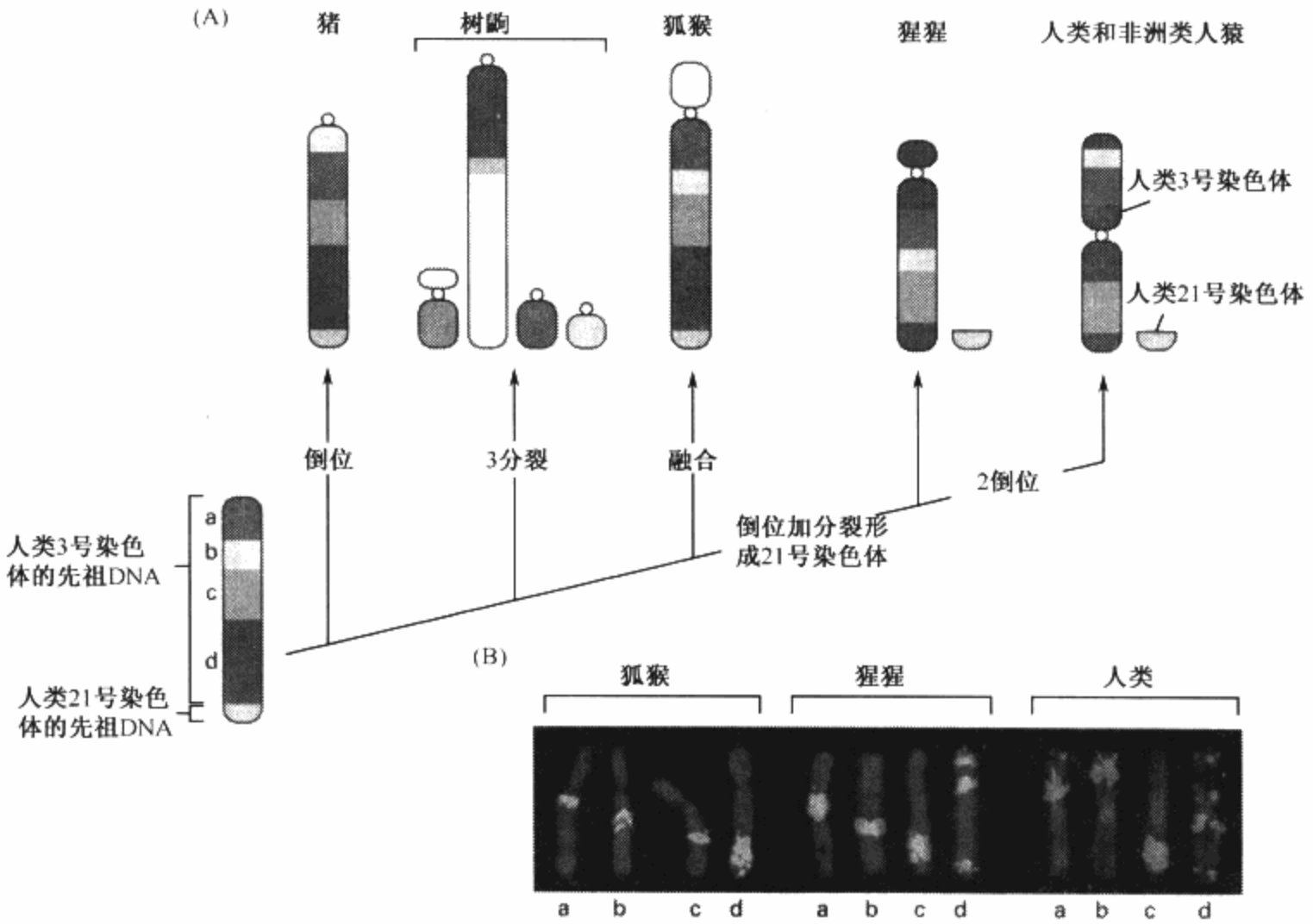


图 4-19　建议中的人类 3 号染色体进化历史和它在其他哺乳动物中的亲属

（A）左下方是假设存在于某个哺乳动物先祖染色体上 3 号染色体片段的顺序。上面排列的是现今哺乳动物染色体中发现的染色体带型。沿每个分支标注的是从假设的先祖得到现今染色体的外形所需的最少变化。在哺乳动物中，认为这些变化类型在染色体组织结构中每 $5\times10^6\sim10\times10^6$ 年发生一次。现今染色体中描绘的小环代表着丝粒的位置。（B）产生（A）中图形的一些染色体染色实验。每个图显示了与人类 3 号染色体亲缘关系很近的染色体，它们与不同的 DNA 片段杂交后染成绿色，并在图的底部标以 a、b、c 和 d 4 个字母。这些字母对应于（A）中图的彩色片段（摘自 S. Müller et al., *Proc. Natl. Acad. Sci. USA* 97：206～211，2000. © Natinal Academy of Sciences）

染色体存在于细胞生命过程中的不同阶段

我们已经看到了基因在染色体上是如何排列的，但是，要形成具有功能的染色体，DNA 分子必须具备更多的功能而不仅仅是携带基因：它必须能够进行复制，且复制后的拷贝必须分离，并在每一次细胞分裂时，可靠地分配到子代细胞。这个过程需经历一系列有序的阶段，

它们一起称为**细胞周期**。图 4-20 简要地总结了细胞的周期，并在第 17 章进行详细的讨论。本章只关注细胞周期中的两个阶段。在**有丝分裂间期**，染色体进行复制，而在**有丝分裂期间**，它们变得高度凝聚，然后发生分离，并分配给两个子细胞核。分裂细胞中存在的这种高度凝聚的染色体称为**有丝分裂染色体**。这是染色体最容易观察到的形式；实际上到目前为止，本章所示的所有染色体图都是处于有丝分裂期的染色体。这种凝聚的状态对于细胞分裂时有丝分裂纺锤体将复制后的染色体分离是很重要的，这些将在第 18 章讨论。

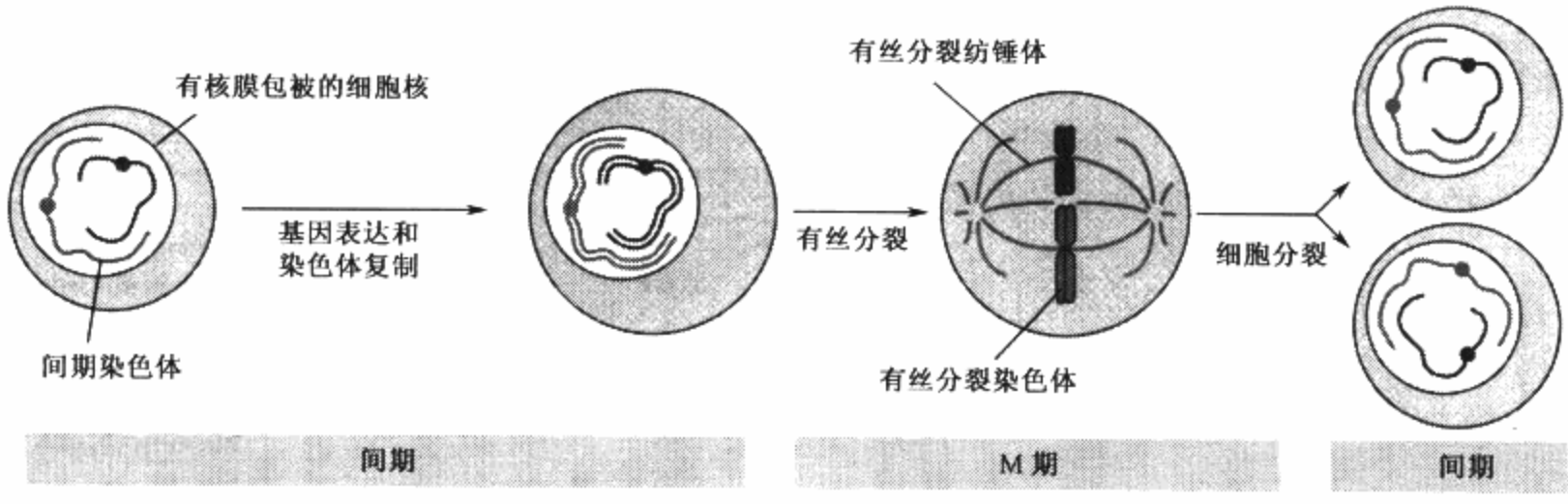


图 4-20　真核细胞周期简图

在有丝分裂间期，细胞活跃地表达着它的基因，因而不断合成蛋白质。而且在间期和细胞分裂之前，DNA 发生复制，染色体也得以加倍，一旦 DNA 复制完成，细胞就可以进入 M 期，在这一阶段发生有丝分裂，细胞核分裂成两个子细胞核。在此期间，染色体凝聚，核膜破裂，微管和其他蛋白形成有丝分裂纺锤体。有丝分裂纺锤体附着凝聚的有丝分裂染色体，然后将一套完整的染色体向细胞两端牵引。核膜围绕每个染色体组重建起来，并在 M 期最后一步，细胞分裂成两个子细胞。细胞周期大部分时间都花费在间期，相比之下 M 期很短暂，在许多哺乳动物中大约只占 1h 的时间

在细胞周期的其他阶段，细胞没有进行分裂时，染色体就会伸展开，而且它们的大部分染色质以细胞核内长而细的缠结存在，这样就不容易分辨出单个的染色体（图 4-21）。我们称这种伸展状态的染色体为有丝分裂间期染色体。

每个形成线性染色体的 DNA 分子都必须含有一个着丝粒、两个端粒和复制起点

每条染色体都作为一个不同结构单位发挥功能：作为一个拷贝为了在细胞分裂时能传递到每个子细胞中，每条染色体必须能够进行复制，随后新复制的染色体必须分离，并正确地分配到两个子细胞中。这些基本的功能由 DNA 中 3 种特殊核苷酸序列进行控制，它们中每一个都与引导染色体复制和分离装置的特殊蛋白质结合（图 4-22）。

酵母的染色体相对较小，并且易于操作，在其中进行的实验鉴定出了负责上述每种功能的最小 DNA 序列元件。其中，一种核苷酸序列的功能是作为 DNA 的**复制起点**，也就是 DNA 开始复制的位置。真核生物染色体含有许多复制起点，以确保整条染色体能迅速地完成复制，这些在第 5 章中详细讨论。

复制完毕以后，两条子染色体仍然彼此附着，并且随着细胞周期的进行，进一步凝聚，形成有丝分裂染色体。另一种称作**着丝粒**的特殊 DNA 序列的存在，使每条复制并凝聚的染色体在细胞分裂时能向各个子细胞牵引。在着丝粒位置形成了一个称作**动粒**的蛋白质复合物，它将复制后的染色体附着到有丝分裂纺锤体上，使它们能够被分开（讨论见第 18 章）。

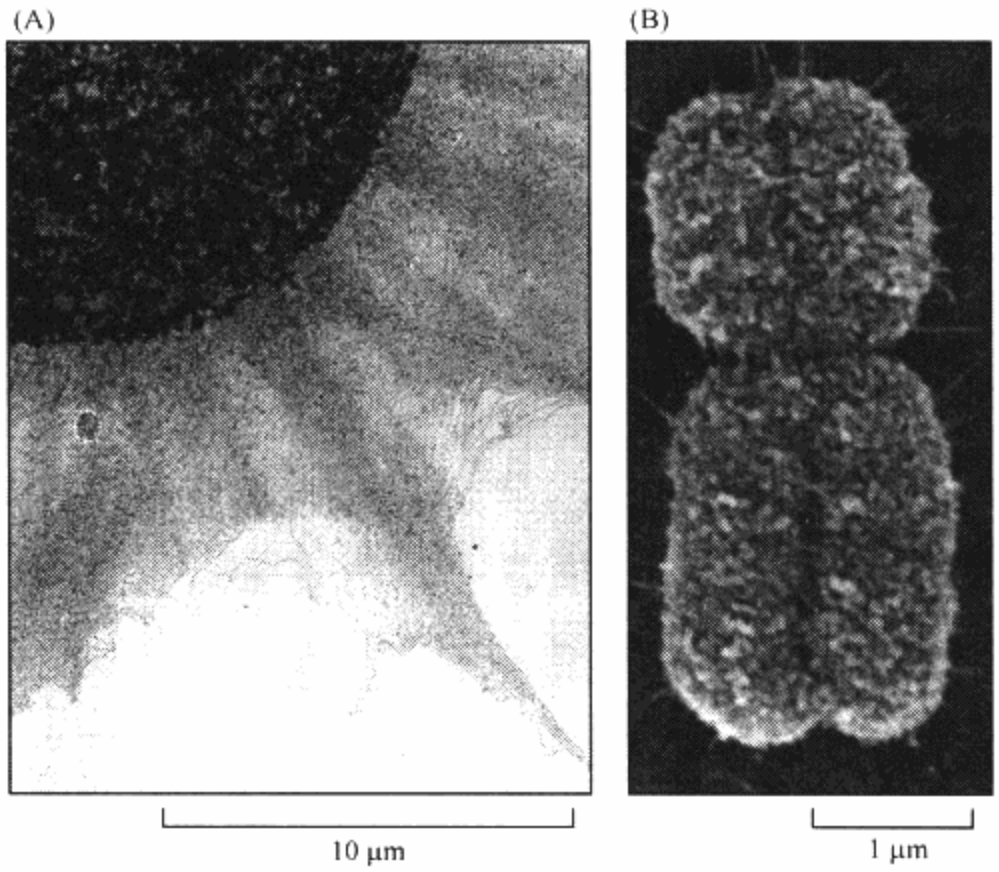


图 4-21　比较伸展的间期染色质与有丝分裂染色体中的染色质

（A）一张电镜显微照片，显示了巨大的染色质缠结从裂解的间期细胞核溢出。（B）一张有丝分裂染色体的扫描电镜照片：一条凝聚且已倍增的染色体，其中的两条新染色体仍然连接在一起（图 4-22）。缢痕区域表示着丝粒位置。注意比例尺的差别[（A）经 Victoria Foe 许可；（B）经 Terry D. Allen. 许可]

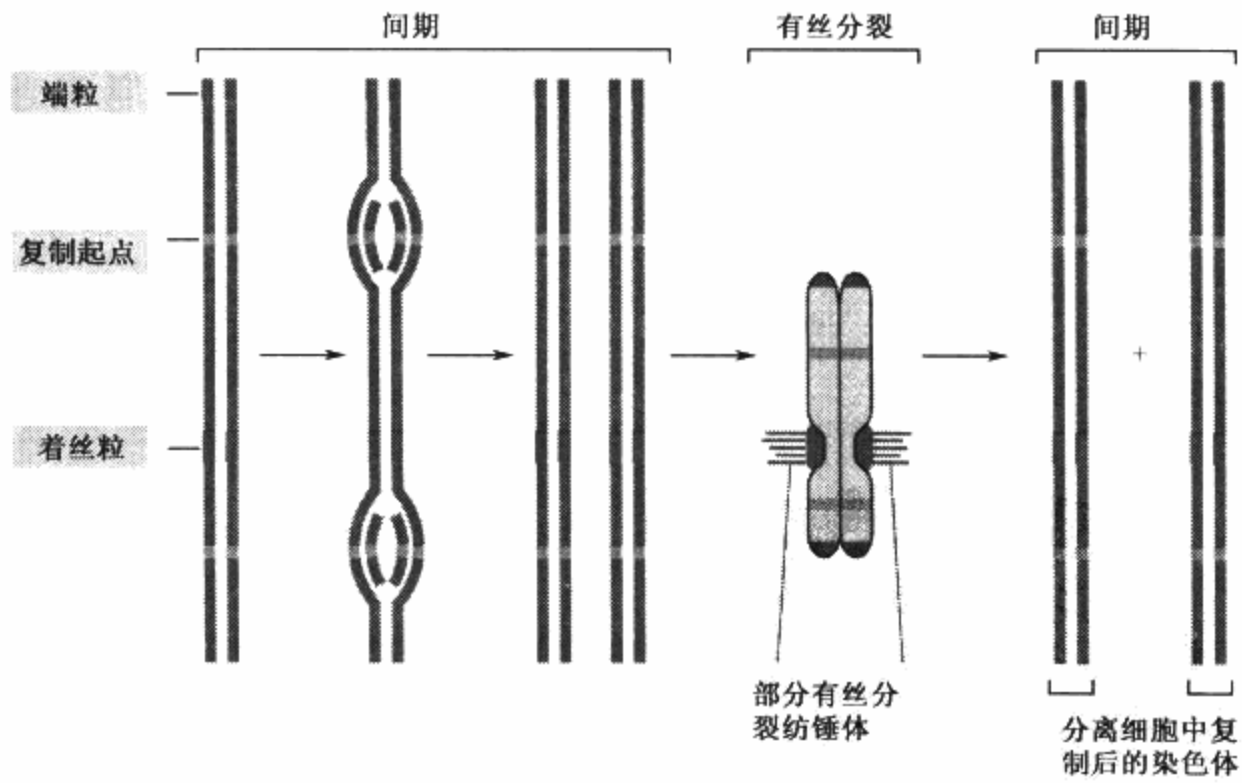


图 4-22　形成一条能够进行复制，且在有丝分裂时能发生分离的真核染色体所需的 3 种 DNA 序列

每条染色体含有多个复制起点、1 个着丝粒和 2 个端粒。这里所示的是一条典型染色体在细胞周期中要发生的事件顺序。在有丝分裂间期 DNA 复制从复制起点开始，并从这些起点双向前进跨越该染色体。在 M 期，着丝粒将加倍的染色体附着到有丝分裂纺锤体上，这样在有丝分裂时就会在每个子细胞中分配一份拷贝，着丝粒也有助于将加倍的染色体维系在一起，直到它们准备好彼此被分离。端粒在每个染色体末端形成特殊的帽子

第三种特殊的 DNA 序列形成**端粒**，也就是染色体的末端。端粒含有重复的核苷酸序列，使染色体末端能够有效地复制。端粒也行使其他功能：重复的端粒 DNA 序列及其相邻区域形成一种结构，能够防止细胞将染色体末端识别为需要修复的断裂 DNA 分子。我们将在第 5 章讨论这种类型的修复以及端粒的其他特征。

在酵母细胞中，染色体增殖所需的这 3 种序列都比较短（通常每种不到 1000 个碱基对），因此仅仅使用了染色体信息携带容量的极小部分。尽管所有真核生物中端粒序列都相当简单且很短，但是规定着丝粒和复制起点的序列在更复杂的生物中比酵母中对应的序列要长得多。例如，实验研究表明人类染色体的着丝粒可能含有多达 100 000 个核苷酸对。甚至有人提出人类着丝粒并不要求存在一段有清晰核苷酸序列界限的 DNA；而仅仅是形成一个巨大的、规律性重复的蛋白质-核酸结构。在本章最后，我们从更一般的方面讨论形成染色体的蛋白质以及 DNA 以后，再回到这个问题上来。

染色体中的 DNA 分子高度凝聚

所有的真核生物都有复杂的方式将 DNA 包装成染色体。回想本章前面讲到过，人类 22 号染色体含有大约 48×10^6 个核苷酸对。它的 DNA 伸展开来，从一端到另一端长达 1.5cm。可是作为有丝分裂染色体存在时，22 号染色体长度仅有 2μm（图 4-10 和图 4-11），这样端到端压缩比达 10 000 倍。这个显著的压缩成果是由许多蛋白质完成的，它们接连地将 DNA 盘绕并折叠成越来越高级的组织形式。间期染色体 DNA 尽管比有丝分裂染色体压缩程度小，但仍然紧密地包装着，其总压缩比大约为 1000 倍。下一部分我们将讨论实现如此压缩程度的那些特殊蛋白质。

阅读本部分内容时，很重要的一点就是要记住染色体的结构是动态变化的，不仅整条染色体对应细胞周期的不同阶段进行着凝聚，而且间期染色体的不同区域，在细胞利用特殊 DNA 序列进行基因表达、DNA 修复以及复制时，能够凝聚和解凝聚。因此。染色体的包装必须以某种方式实现，以便一经要求就能够迅速地局部接触 DNA 序列。

核小体是真核生物染色体结构的基本单位

传统上将结合 DNA 并形成真核生物染色体的蛋白质分为两种一般类型：**组蛋白**和非染色体组蛋白，这两种蛋白质与真核细胞核内 DNA 形成的复合物称为**染色质**。细胞中的组蛋白数量如此之大（每个人类细胞每种组蛋白含量大约为 60×10^6 个分子），以至于染色质中它们的总重量与 DNA 相当。

组蛋白负责形成第一级，也是最基本水平的染色体组织结构——**核小体**，它于 1974 年被发现。将处于有丝分裂间期的细胞核小心地破裂后，在电子显微镜下进行观察它们的成分，发现核内大部分染色体以直径约 30nm 的纤维形式存在[图4-23(A)]。如果对这种染色质进行处理使其部分解折叠，那么在电子显微镜下就可以看到它以许多串在线上的“珠子”形式存在[图4-23(B)]。这条线就是 DNA，而每个珠子是一个“核小体核心颗粒”，它是由 DNA 盘绕在组蛋白形成的蛋白质核心上而形成。串在线上的珠子代表染色体 DNA 包装的一级水平。

最早利用有些能够在核小体之间切断 DNA 的特殊酶（称作核酸酶），消化解折叠的染色体后，分离出核小体后得以确定核小体的组织结构。用这种酶消化较短时间以后，就将暴露在核小体核心颗粒之间的 DNA——**连接** DNA，降解掉。单个的核小体核心颗粒由 8 个组蛋白形成的复合体（组蛋白 H2A、H2B、H3、H4 各有两分子）和长 146 个核苷酸对的双链

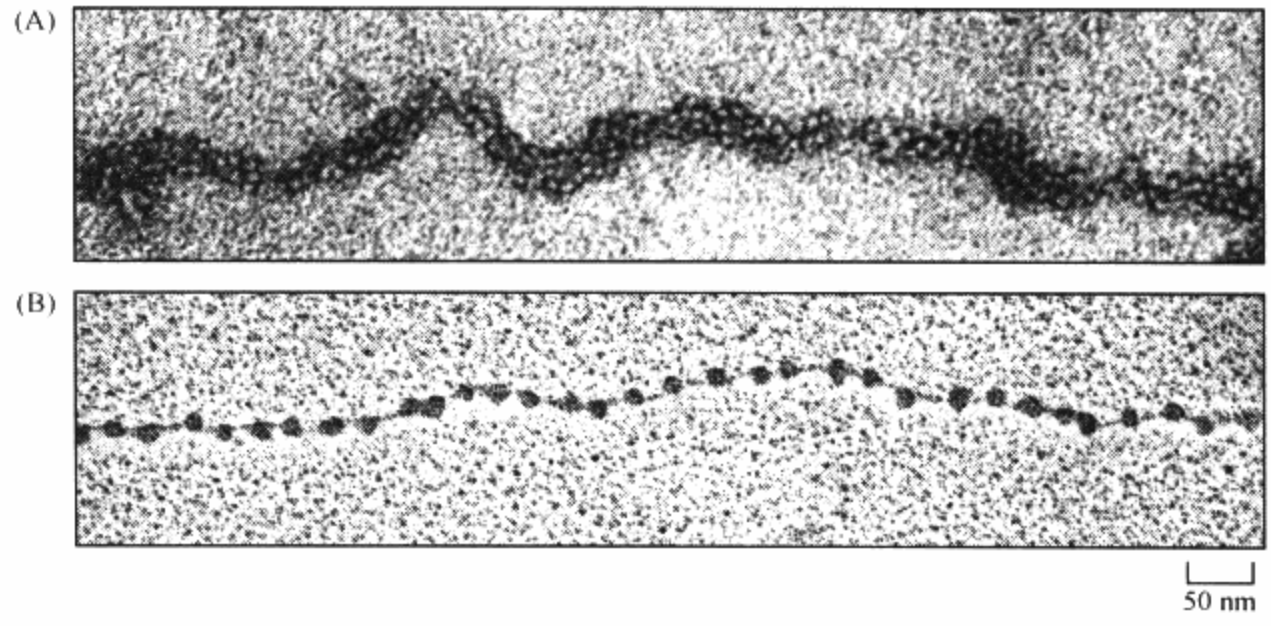


图 4-23　电子显微镜中观察到的核小体

（A）直接从间期细胞核中分离得到的染色质在电子显微镜中呈现出一条厚 30nm 的细线。（B）该电镜照片显示了一段染色质，已用实验方法去包装，或者解凝聚后以便裸露出核小体的[（A）经 Barbara Hamkalo 许可；（B）经 Victoria Foe 许可]

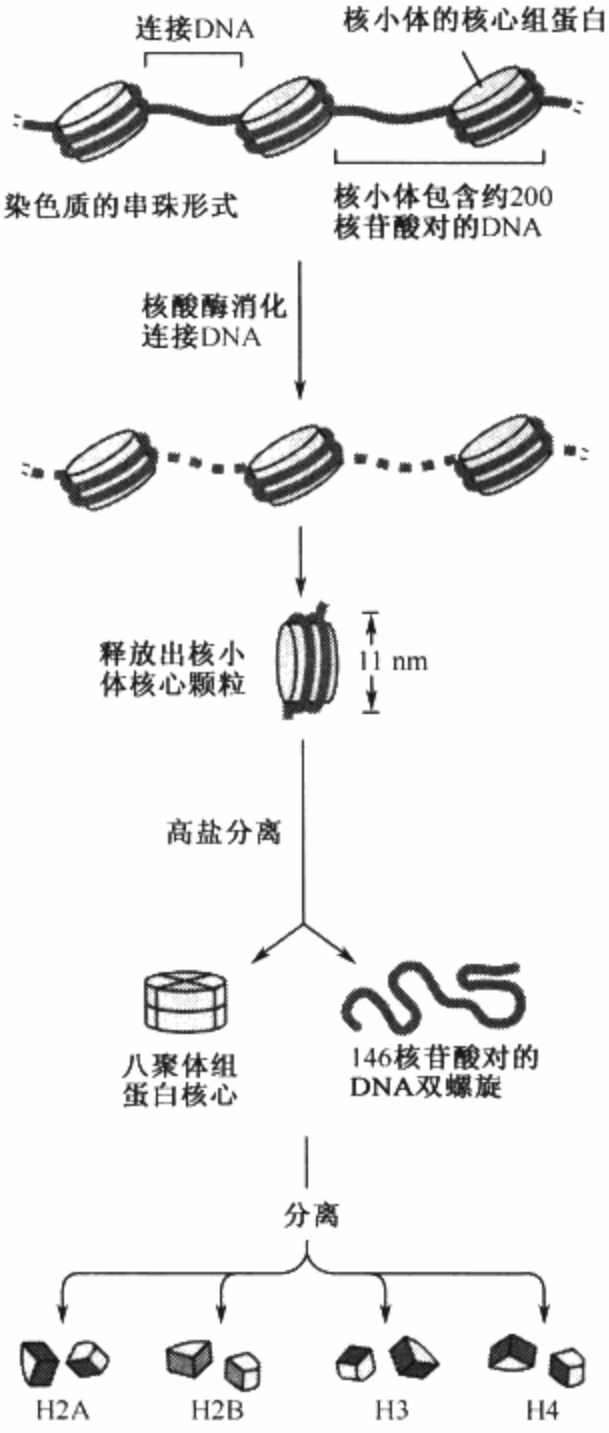


图 4-24　核小体的结构组织

每个核小体含有一个由 8 种组蛋白分子形成的蛋白核心。如图所示，用核酸酶（一种断裂 DNA 的酶）消化连接 DNA 后，从染色体中释放出核小体核心颗粒（该核酸酶能降解裸露出来的连接 DNA，但不能攻击紧密缠绕在核小体核心四周的 DNA）。分离得到的核小体解离成蛋白核心和 DNA 以后，就能确定盘绕核心的 DNA 长度。146 个核苷酸对的长度足以缠绕组蛋白核心 1.65 圈

DNA 组成。这个组**蛋白八聚体**形成了一个蛋白质核心，双链 DNA 则缠绕其上（图 4-24）。

每个核小体核心颗粒之间由一段连接 DNA 区域分隔开，它们长度不一，在几个核苷酸对到 80 个核苷酸对之间变动。（**核小体**这个词按说应该指核小体的核心颗粒及其邻近的一段连接 DNA，但它常用作核小体核心颗粒的同义词。）因此，平均起来核小体间隔大约 200 个核苷酸对进行重复。例如，一个含有 6.4×10^9 核苷酸对的二倍体人类细胞，含有大约 30×10^6 个核小体。核小体的形成将一条 DNA 分子转变成一条染色质线，长度大约为原来的 1/3，这是 DNA 包装的一级水平。

核小体核心颗粒的结构揭示了 DNA 是如何包装的

1997 年得到的核小体核心颗粒高分辨率的结构显示，DNA 以左手螺旋绕圆盘状的组蛋白核心，缠绕 1.65 圈（图 4-25）。形成核小体核心的 4 种组蛋白全部都是比较小的蛋白质（102～135 个氨基酸），它们共有一个相同的结构域，称为组**蛋白折叠**，由 2 个环连接 3 个 α 螺旋而成（图 4-26）。在装配核小体时，首先由组蛋白折叠彼此结合形成 H3-H4 二聚体和 H2A-H2B 二聚体，然后 2 个 H3-H4 二聚体再结合成四聚体，进一步由一个 H3-H4 四聚体结合 2 个 H2A-H2B 二聚体形成致密的八聚体，DNA 就缠绕在它上面（图 4-27）。

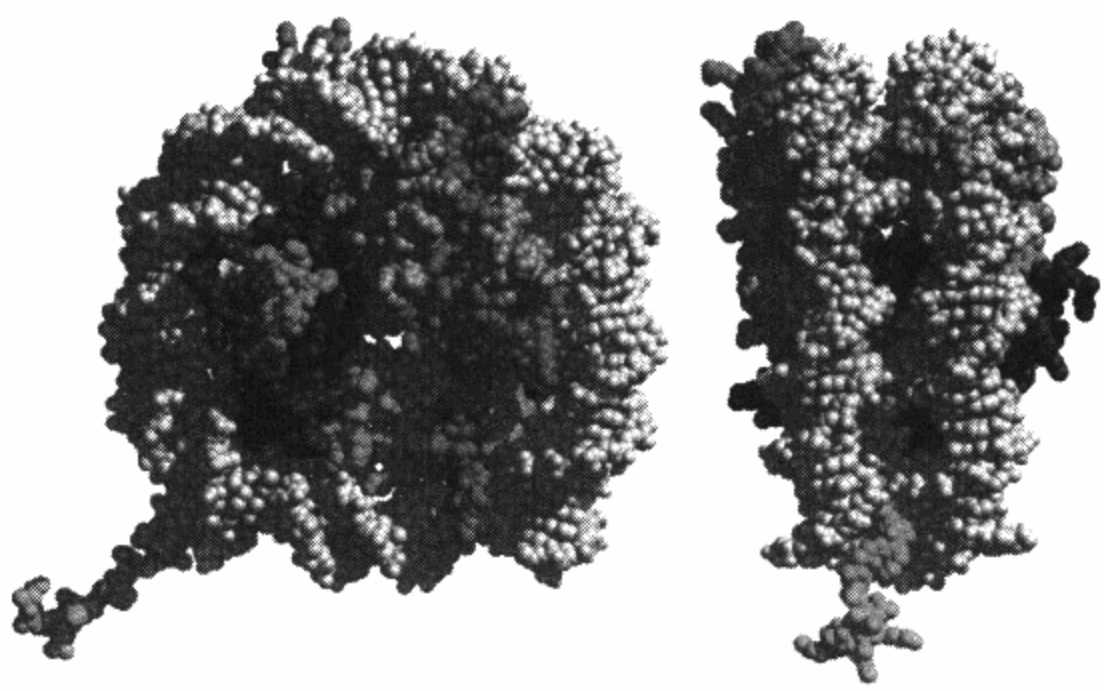


图 4-25　利用晶体 X 射线衍射分析确定的核小体核心颗粒的结构

根据图 4-26 的布局对每个组蛋白着色，淡灰色表示 DNA 双螺旋（摘自 K. Luger et al., *Nature* 389: 251～260, 1997. © Macmillan Magazines Ltd.）

DNA 和组蛋白之间的相互作用很广泛：每个核小体中 DNA 和组蛋白核心之间形成 142 个氢键。这些氢键中将近一半在组蛋白的氨基酸骨架和 DNA 的磷酸二酯骨架之间形成。大量的疏水相互作用和盐键也将核小体中的 DNA 和蛋白质维系在一起。例如，所有的核心组蛋白都富含赖氨酸和精氨酸（2 个拥有基本侧链的氨基酸），它们的正电荷能有效地中和 DNA 骨架所带负电荷，这些大量的相互作用部分解释了为什么实际上任何 DNA 序列都能结合到组蛋白八聚体核心上。DNA 缠绕组蛋白核心的路径并不平坦；相反，在 DNA 上可看到多个结节，从核心不一致的表面也可想到这一点。

每个核心组蛋白除了组蛋白折叠之外，还有一个长长的 N 端氨基酸“尾”，它从 DNA 组蛋白核心延伸出来（图 4-27）。这些组蛋白尾上发生许多不同类型的共价修饰，由它们调控染色体结构的许多方面。我们将在本章的后面讨论这些问题。

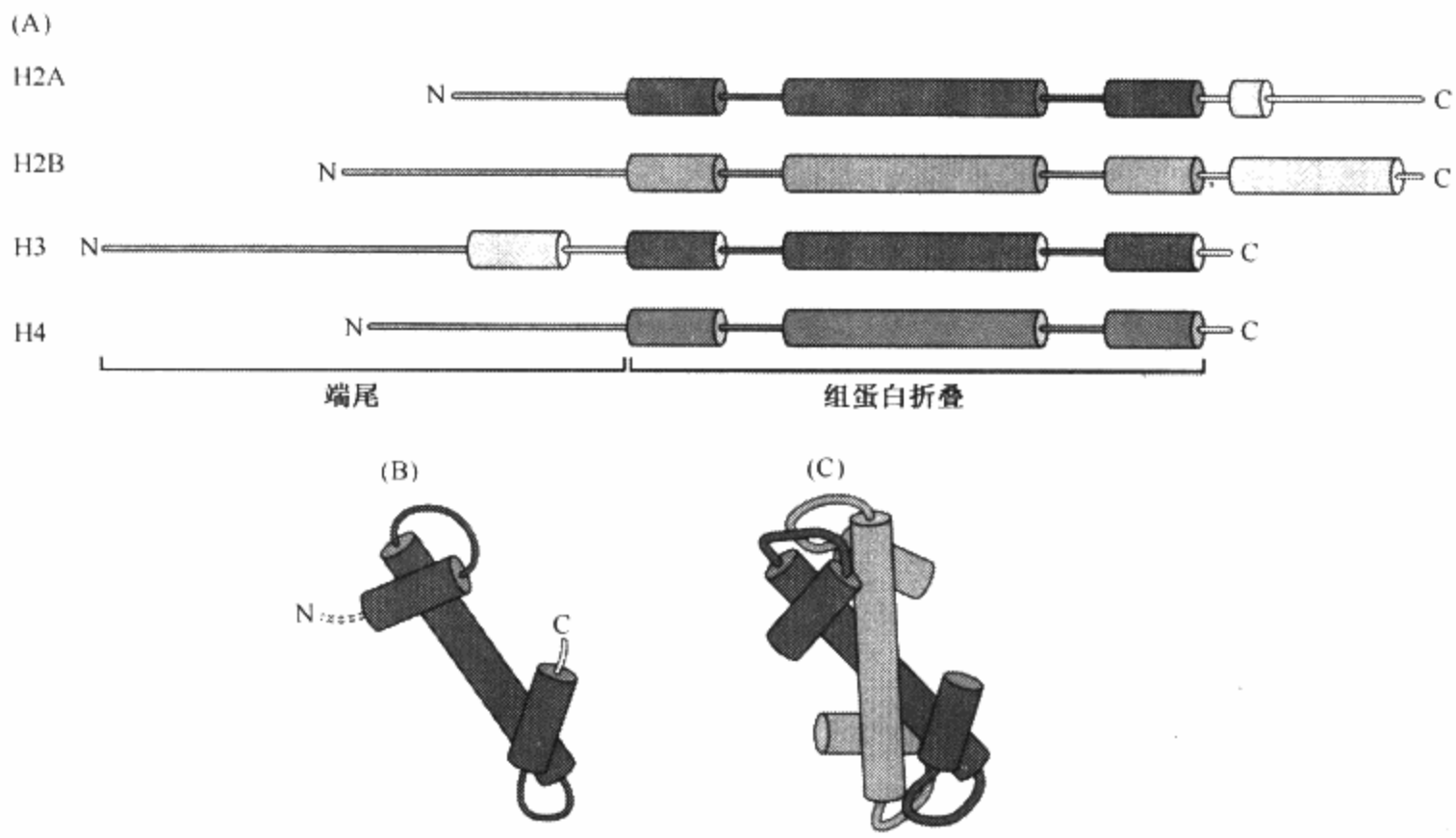


图 4-26　核心组蛋白的整体结构组织

（A）每个核心组蛋白含有 1 个 N 端尾，其上发生多种形式的共价修饰，以及 1 个组蛋白折叠区域，如图所示。（B）组蛋白折叠的结构，它由全部 4 种核心组蛋白形成。（C）组蛋白 H2A 和 H2B 通过一种称为“握手”的相互作用形成一个二聚体。H3 和 H4 以相同的相互作用方式形成一个二聚体，如图 4-27 所示

由组蛋白在 DNA 包装过程中所起的重要作用，可能会推想到它们属于最保守的真核生物蛋白质之列。例如，来自于豌豆和牛的组蛋白 H4 的氨基酸序列 102 个位点中仅有 2 个不同，这种极强的进化保守性表明组蛋白功能几乎涉及所有的氨基酸，所以任何位置的改变对于细胞都是有害的。在酵母细胞中，可以通过**体外**突变给定的组蛋白基因，然后将其引入到酵母基因组中取代原来正常的基因，从而直接验证这一点。组蛋白基因序列大部分改变都是致命的；即便少数不足以致命，也会引起基因表达正常模式的改变以及其他方面的异常。

尽管核心组蛋白高度保守，许多真核生物也合成各种特殊的核心组蛋白，它们的氨基酸序列与主要的核心组蛋白不同。例如，海胆有 5 种组蛋白 H2A 变体，它们每一种都在发育的不同时期表达。有人认为核小体已经纳入了这些稳定性与常规核小体不同的变体组蛋白，并且它们可能特别适合发育早期阶段迅速的 DNA 转录和 DNA 复制。

核小体在 DNA 上的位置由 DNA 柔性和其他 DNA 结合蛋白质决定

尽管从理论上讲，几乎每一种 DNA 序列都能折叠成核小体，但是细胞中核小体间隔可能不规则，两个主要影响因素决定了在 DNA 中什么位置能形成核小体。一个就是双螺旋 DNA 弯曲形成绕组蛋白八聚体外围的两圈紧密圆环的难度，因为这个过程大量地压缩 DNA 螺旋的小沟。由于小沟中富含的 A-T 序列比 G-C 序列更容易压缩，每个组蛋白八聚体倾向将自身定位于 DNA 上某个位置以便让处于 DNA 螺旋内侧的小沟最大化（图 4-28）。因此一段 DNA 片段，如果富含 A-T 的短序列，并由整数倍的 DNA 环间隔开，比起那些缺乏这些特征的 DNA 片段，就更容易绕核小体弯曲。此外，由于核小体中的 DNA 在许多地方形成结节，所以给定的一段核苷酸能够容纳这种变形的能力，也能影响 DNA 在核小体

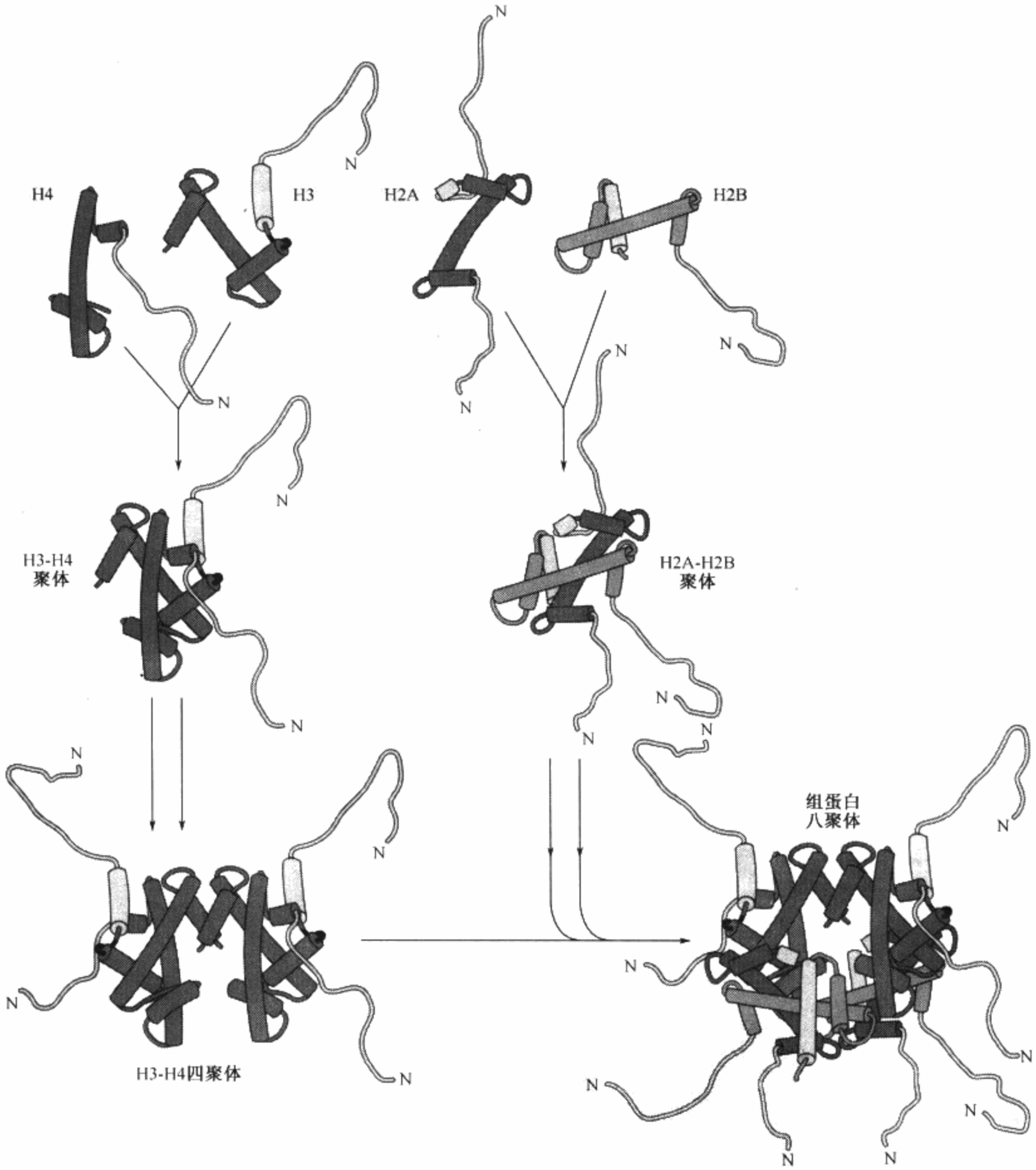


图 4-27　组蛋白八聚体的装配

组蛋白 H3-H4 二聚体和 H2A-H2B 二聚体通过握手相互作用形成。一个 H3-H4 四聚体形成八聚体骨架，然后 2 个 H2A-H2B 二聚体结合上来，完成组装过程。组蛋白按图 4-26 中那样着色。注意这些组蛋白全部 8 个 N 端尾都从圆盘形核心结构上伸出。在 X 射线晶体结构中（图 4-25），大部分组蛋白尾处于无组织状态（并因此在结构中看不到），表明它们的构象是有高度柔性的（摘自 J. Waterborg 绘制的图）

上的位置。

DNA 的这些特征可能也解释了某些沿 DNA 进行非常精确地定位，显著却又少见的情况。然而对于大部分染色体中的 DNA 序列，并不存在与核小体强烈优先结合的核小体结合位点；核小体可以占据相关 DNA 序列上许多位点中的任何一个。

第二个，也可能是最重要的影响核小体定位的因素是 DNA 上其他牢固结合蛋白质的存在。一些结合蛋白质喜欢在它们附近形成核小体，另一些则形成障碍，强行将核小体定位在它们之间的位置。最后，还有些蛋白质能牢固地结合到 DNA 上，即使它们的 DNA 结合位点是核小体的一部分。因此核小体在一段 DNA 上的确切位置由几个因素决定，包括 DNA 序列和是否存在结合到 DNA 上的蛋白质及其特性。而且，下面我们将看到，核小体在 DNA 上的排列是高度动态的，根据细胞的需要在迅速地发生变化。

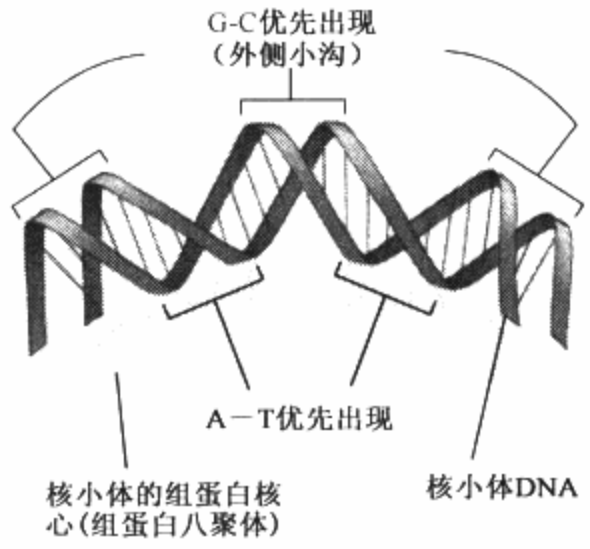


图 4-28　核小体中 DNA 的弯曲

DNA 双螺旋围绕组蛋白八聚体紧密缠绕 1.65 圈。这幅图以大致的比例绘制，显示了在转角内侧如何压缩小沟。由于 DNA 分子的某些结构特征，A-T 碱基配对优先出现在这种狭窄的小沟中

核小体通常一起包装成一条致密的染色质纤维

尽管大多数染色体 DNA 上形成了长串的核小体，但活细胞中的染色质可能基本不采用这种伸展的“串珠状链”的形式。相反，核小体一个叠一个地组装起来，形成一种规则的排列，甚至进一步压缩了其中的 DNA。因此将细胞核小心地裂解到电子显微镜的载网上的时候，可以看到大部分染色质以直径大约 30nm 的纤维形式存在，这比以“串珠状链”形式存在的染色体要宽得多（图 4-23）。

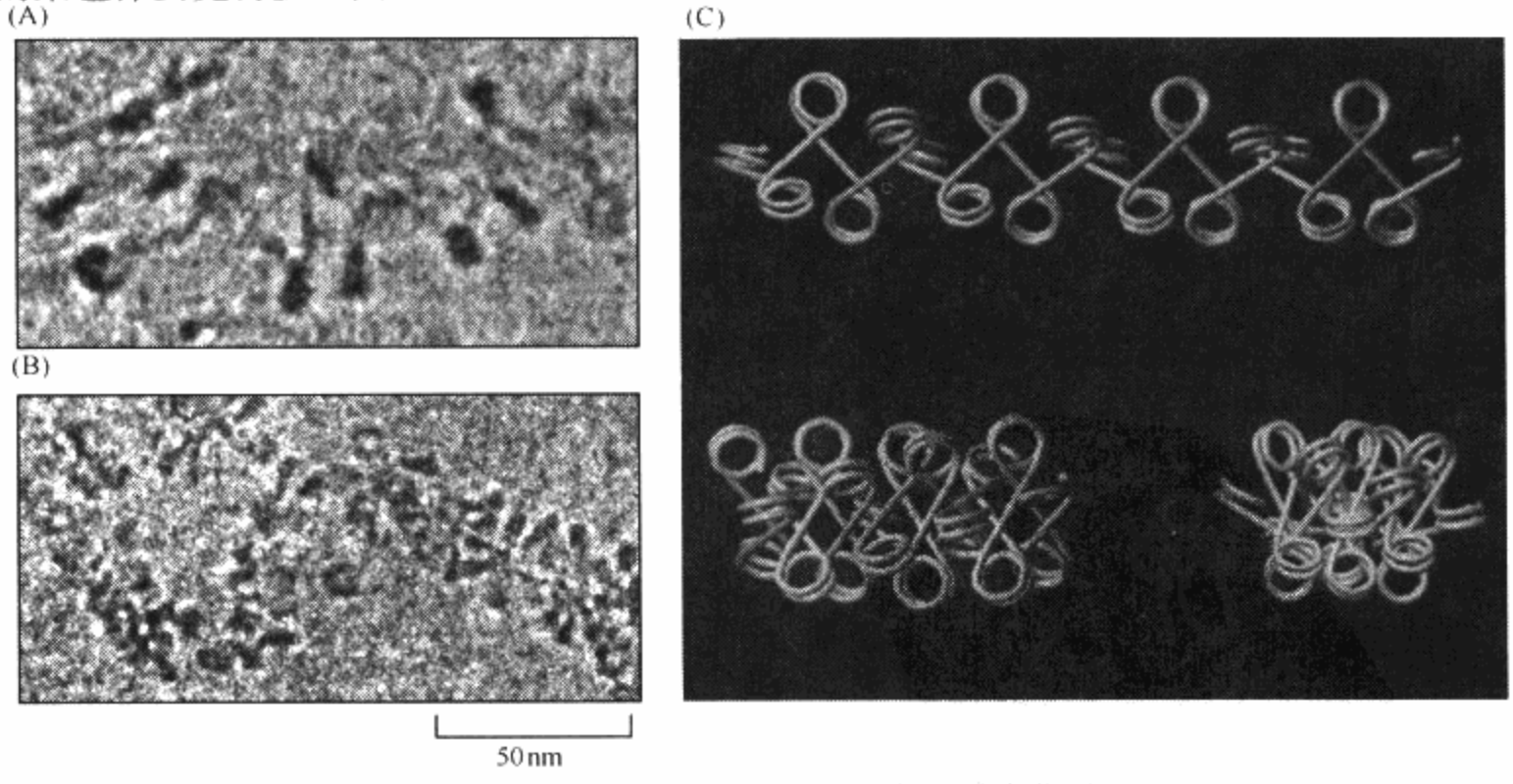


图 4-29　30nm 染色质纤维 Z 字形模型的变化形式

（A）和（B）支持存在（C）中所描绘的右上和右下模型结构的电子显微照片证据。（C）Z 字形变化形式。有人提出通过纤维长度类似手风琴的伸展和压缩来实现这三种变化之间的相互转换。连接 DNA 的曲折前进或缠绕，或者纤维宽度的局部微小变动，可以容纳相邻核小体珠子之间连接 DNA 长度的差异。30nm 纤维的形成既需要组蛋白 H1，也需要核心组蛋白尾；为简便起见，这里两者都未示出，但可见图 4-30 和图 4-32（摘自 J. Bednar et al., *Proc. Natl. Acad. Sci. USA* 95：14173～14178，1998. © National Academy of Science）

已经提出多种模型用于解释核小体在 30nm 染色质纤维中是如何包装的；但其中与能够获得的数据最一致的模型是一系列的结构变化，它们统称为 Z 字形模型（图 4-29）。事实上，染色体中发现的 30nm 结构可能是不同的 Z 字形的变化形式镶嵌而成的一种易变的模式。前面我们看到连接相邻两个核小体的连接 DNA 长度可以不同，连接 DNA 长度的这些差别可能进一步扰乱了局部的 Z 字形结构。最后，其他 DNA 结合蛋白质和难以折叠成核小体的 DNA 序列的存在也使 30nm 纤维间断出现了不规则的特征（图 4-30）。

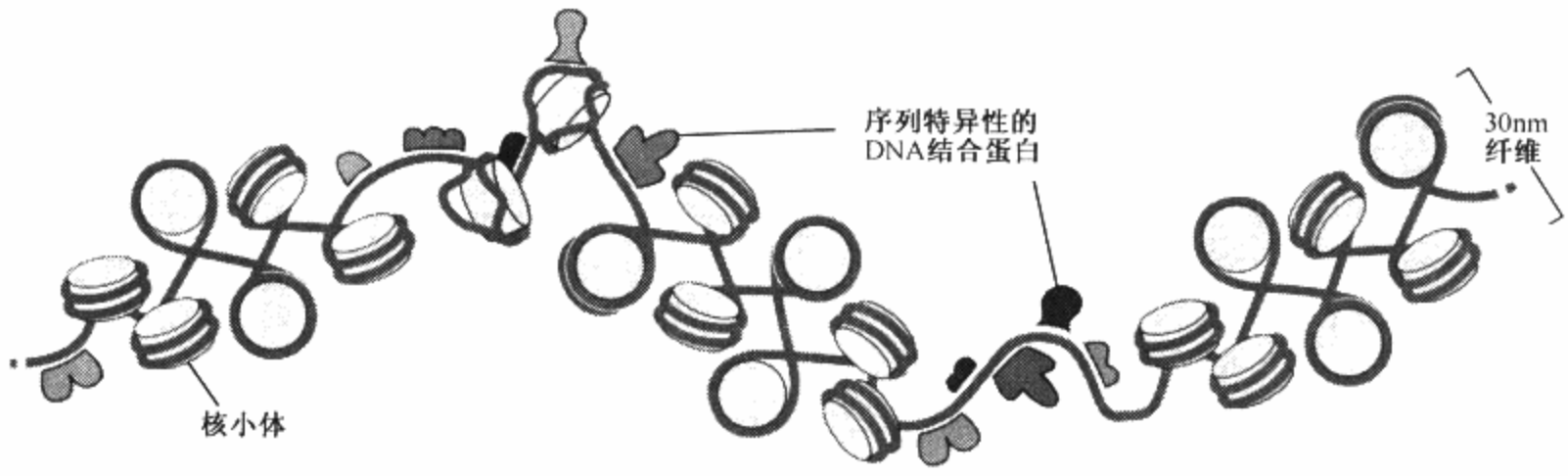


图 4-30　30nm 纤维中的不规则性

这幅 30nm 纤维的简图说明了序列特异的 DNA 结合蛋白质对其结构的阻断。第 7 章解释了这些蛋白质是怎样牢固结合 DNA 的。30nm 纤维的阻断可能是因为存在完全缺少核小体的 DNA 区域，或者更可能是因为存在含有改变过或重建过的核小体的区域。通过实验的方法常常发现，不含核小体或含有重建核小体区域的 DNA 对核酸酶消化异常敏感（与核小体中的 DNA 相比）

多种机制可能一起发挥作用，从线状的核小体链形成 30nm 纤维。首先，另外一种组蛋白，称作组蛋白 H1，参与了此过程。H1 比核心组蛋白更大，而且远远不及核心组蛋白保守。实际上，大部分真核生物的细胞合成许多组蛋白 H1，它们密切相关，但具有相当不同氨基酸序列。

每个核小体结合一个组蛋白 H1 分子，从而将 DNA 和蛋白质联系起来，并在 DNA 从核小体出来时，改变它的延伸路径。至于 H1 是如何将多个核小体一起包装进 30nm 纤维，其细节尽管还不清楚，但是看起来 DNA 伸出路径的改变对于压缩核小体 DNA 似乎起着关键作用，因为这样它就可以彼此连锁起来形成 30nm 纤维（图 4-31）。

另外一种形成 30nm 纤维的机制可能涉及从核小体伸出的核心组蛋白尾，这在前面我们

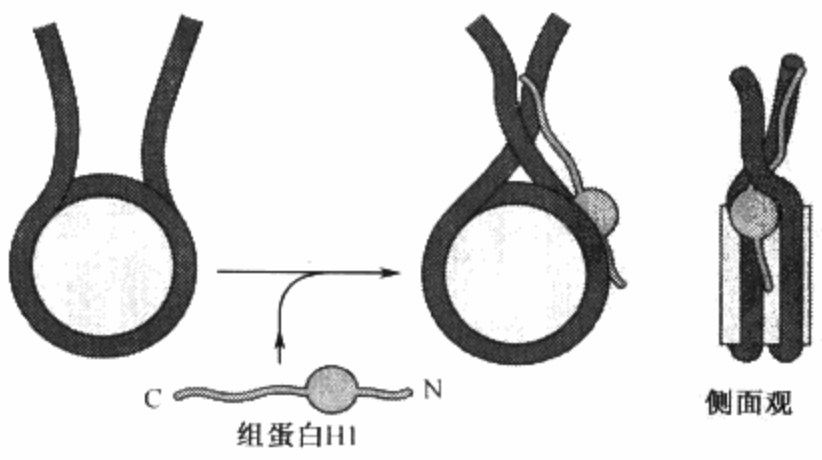


图 4-31　DNA 从核小体出来时组蛋白 H1 怎样改变其路径的猜测模型

组蛋白 H1（绿色）由 1 个球状核心和 2 个伸展的尾组成。H1 对核小体结构压缩的作用一部分可能来自电荷的中和；与核心组蛋白一样，H1 带正电荷（尤其是它的 C 端尾），这有助于压缩带负电荷的 DNA。与核心组蛋白不同的是，H1 似乎对细胞存活不是必需的；在一种纤毛原虫中，缺少组蛋白 H1 的细胞核增加了将近 2 倍，但细胞看起来是正常的

已经看到。有人认为这些尾可能有助于核小体彼此附着——因而在组蛋白 H1 的帮助下将一串核小体压缩成 30nm 纤维（图 4-32）。

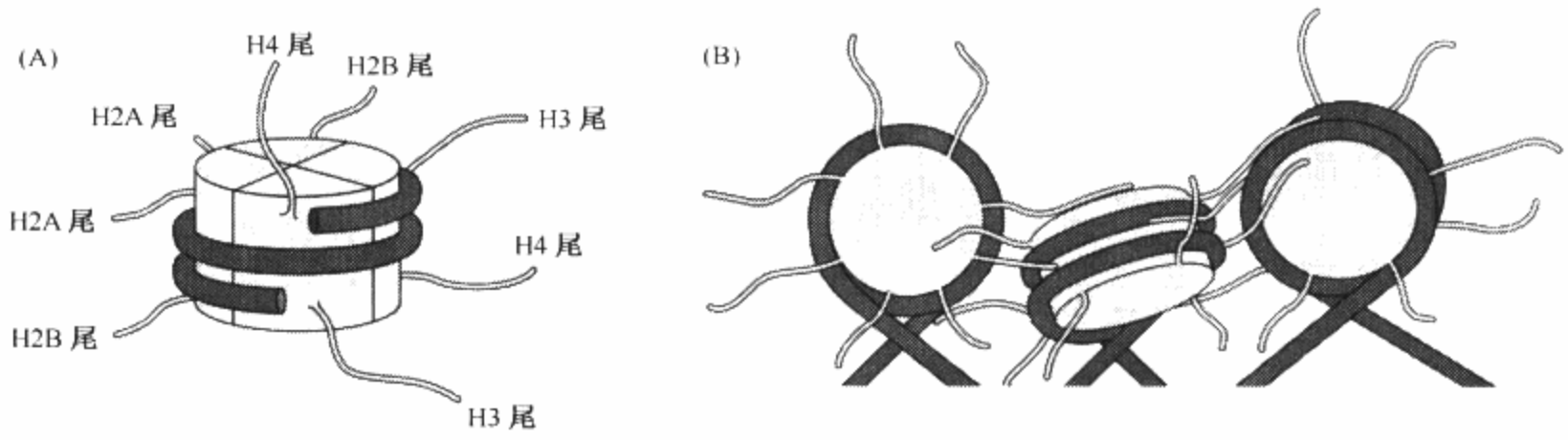


图 4-32　形成 30nm 纤维的组蛋白尾猜测模型

（A）从每个核小体伸出的 8 条组蛋白尾的大致出口位点，其中每个组蛋白亚基伸出 4 条。在核小体高分辨率结构中（图 4-25），这些尾非常缺少组织性，表明它们具有高度的柔性。（B）一个猜测模型，显示了组蛋白尾可能是如何帮助核小体包装形成 30nm 纤维的。该模型基于：①表明组蛋白尾帮助形成 30nm 纤维的证据；②核小体 X 射线衍射晶体结构，它显示在晶格中一个核小体的尾与相邻核小体组蛋白核心接触；③表明组蛋白尾与 DNA 相互作用的证据

ATP 驱动的染色质重建装置改变了核小体结构

多年以来生物学认为，一旦在 DNA 某个特定位置形成一个核小体，由于存在核心组蛋白和 DNA 之间牢固的结合，它就会在这个位置固定不动。但是，最近发现真核细胞含有**染色质重建复合体**，它们是一种蛋白质装置，利用 ATP 水解释放的能量暂时改变核小体的结构，这样 DNA 对组蛋白核心的结合就变得比较松弛。重建后的状态可能主要因为 H2A-H2B 二聚体在核小体核心中发生移动而形成；H3-H4 四聚体特别稳定，所以难以重新排列（图 4-27）。

核小体结构的重建产生两个重要的结果。首先它使细胞内其他蛋白质能迅速接触核小体 DNA，尤其是那些参与基因表达、DNA 复制以及修复的蛋白质。即便在这种重建复合体解体以后，核小体仍然保持处于“重建状态”，含有 DNA 和全部的结合组蛋白，只不过核小体中 DNA-组蛋白结合变得松弛罢了；然后这种重建状态缓慢地恢复到重建前的标准核小体状态。其次，重建复合体能够催化 DNA 上的核小体位置发生改变（图 4-33）；有些甚至能将组蛋白核心从一条 DNA 分子转移到另一条 DNA 分子。

细胞有多种不同的染色质重建复合体，它们的特性存在细微的差别。其中大部分是巨大的蛋白质复合体，含有超过十多个亚基。可能无论什么时候真核细胞需要直接接触核小体 DNA，以便进行基因表达，DNA 复制或者 DNA 修复时，都会利用到它们。不同的重建复合体可能具有专门针对其所起作用的特征。有人认为一部分重建复合体主要作用就是允许蛋白质接触核小体 DNA，而另一部分重建复合体则在不要求接触核小体 DNA 的时候重新形成核小体（图 4-34）。

细胞小心地调控着染色质重建复合体。我们将在第 7 章看到，当基因打开和关闭时，这些复合体会被带至 DNA 的特殊区域，在这里它们发挥作用，影响局部染色质结构。在有丝分裂期间，至少有些染色质重建复合体因磷酸化而失活，这样可能有助于紧密包装的有丝分裂染色体维持其结构。

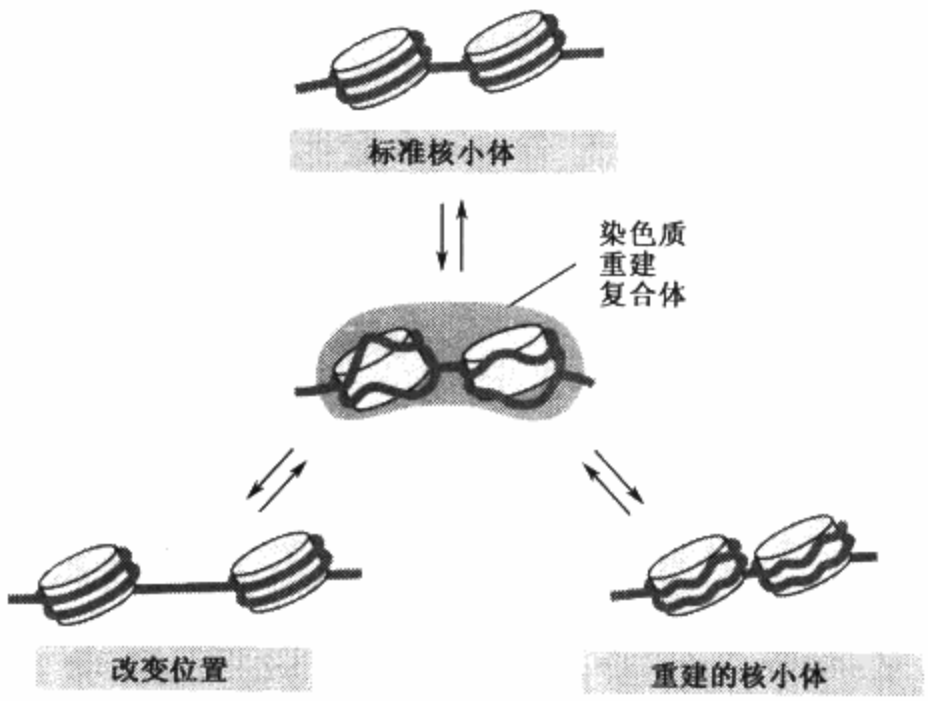


图 4-33　解释某些重建复合体机制的模型

不存在重建复合体时，由于较高的活化能障碍，所示的三种核小体状态之间的相互转换非常缓慢。染色质重建复合体（绿色）利用 ATP 水解释放的能量，形成一个活化的中间产物（中间图所示），其中的组蛋白-DNA 接触受到部分阻断。然后这种活化的状态可以衰退成所示三种核小体构型中的任何一种。通过这种方式，重建复合体大大增加了不同核小体状态之间相互转换的速率。DNA-组蛋白接触发生松弛的重建状态，是一个比标准核小体有更高自由能水平的状态，并将缓慢逆转成标准核小体构象，即便缺少重建复合体也如此。细胞拥有许多不同的染色质重建复合体，它们在具体的生物化学特性方面存在差别。例如，不是所有重建复合体都能改变一个核小体的位置，但是所有的都利用 ATP 水解改变核小体结构（摘自 R. E. Kingston and G. J. Narlikar，*Genes Dev*，13：2339～2352，1999）

组蛋白尾的共价修饰可以对染色质产生深远的影响

4 种核心组蛋白中每一个的 N 端尾在序列上都高度保守，并在调节染色质结构中发挥关键作用，每一条尾都经过多种类型的共价修饰，包括赖氨酸的乙酰化、赖氨酸的甲基化和丝氨酸的磷酸化[图4-35(A)]。组蛋白在细胞质内合成，然后组装成核小体。组蛋白尾修饰有些就发生在刚合成之后，但在组装之前。可是我们关注的是一旦核小体装配起来以后所发生的修饰。这些核小体的修饰与去修饰都由细胞核内的酶催化进行。例如，组蛋白乙酰基转移酶（HAT）催化在组蛋白尾上增加一个乙酰基，而去掉乙酰基则由组蛋白脱乙酰酶（HDAC）催化。

组蛋白尾的不同修饰，产生许多重要的结果。尽管尾的共价修饰对单个核小体的稳定性的直接影响很小，但是这种修饰似乎会影响 30nm 染色体纤维以及下面讨论的更高级结构的稳定性。例如，组蛋白乙酰化会降低染色质结构的稳定性，部分可能是因为增加一个乙酰基后就从赖氨酸上移走了正电荷，因而在压缩染色质时，组蛋白要中和 DNA 上的负电荷就更加困难。然而，修饰后的组蛋白尾产生的最深远的影响就是，它们能够吸引特殊蛋白质结合一段经过适当修饰的染色质。根据组蛋白尾的精确修饰情况，这些蛋白质能够促使染色质进一步压缩，或者能够促进与 DNA 的接触。如果考虑到修饰之间的组合情况，每条组蛋白尾可能存在的不同修饰标记就是一个十分巨大的数目了。因此，有人提出，给定的一段染色质通过组蛋白尾共价修饰，能够传递给细胞具有特定意义的信息[图4-35(B)]。例如，可能某种标记传递的信息表示一段染色质已完成复制，而另一种标记则表示不应该进行基因表达。根据这个观点，每种不同的标记将会吸引那些将随后执行相应功能的蛋白质。由于组蛋白尾是伸展的，并因而即便在染色质处于凝聚状态时，也可以接触到，所以它们为这类信息提供了

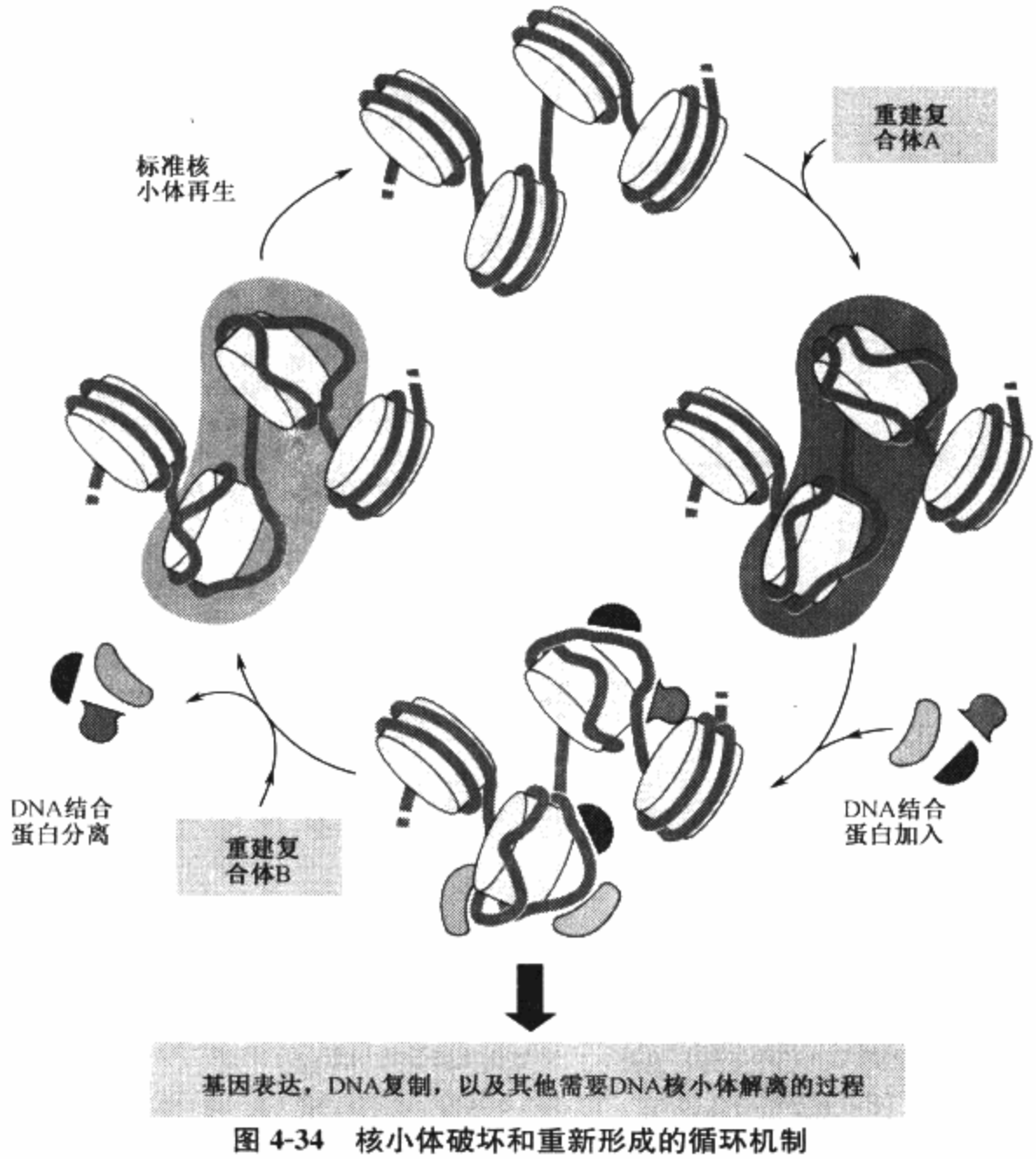
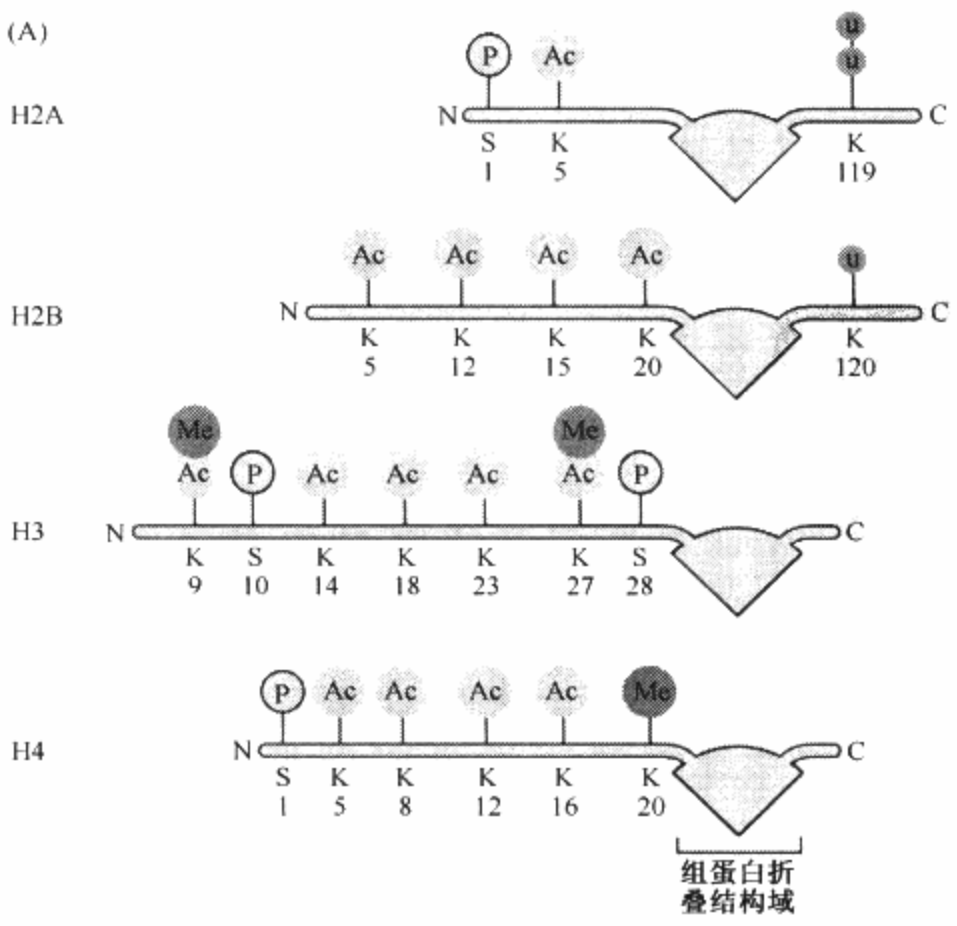


图 4-34　核小体破坏和重新形成的循环机制

根据这个模型，尽管从理论上讲，相同复合体可能催化两个反应，但不同的染色质重建复合体破坏并重新形成核小体。DNA 结合蛋白可以在基因表达、DNA 复制或 DNA 修复中发挥作用，并在有些情况下，它们结合在一起可以引起组蛋白核心解离，以便形成像图 4-30 所示的无核小体 DNA 区域（摘自 A. Travers，*Cell* 96：311～314，1999）



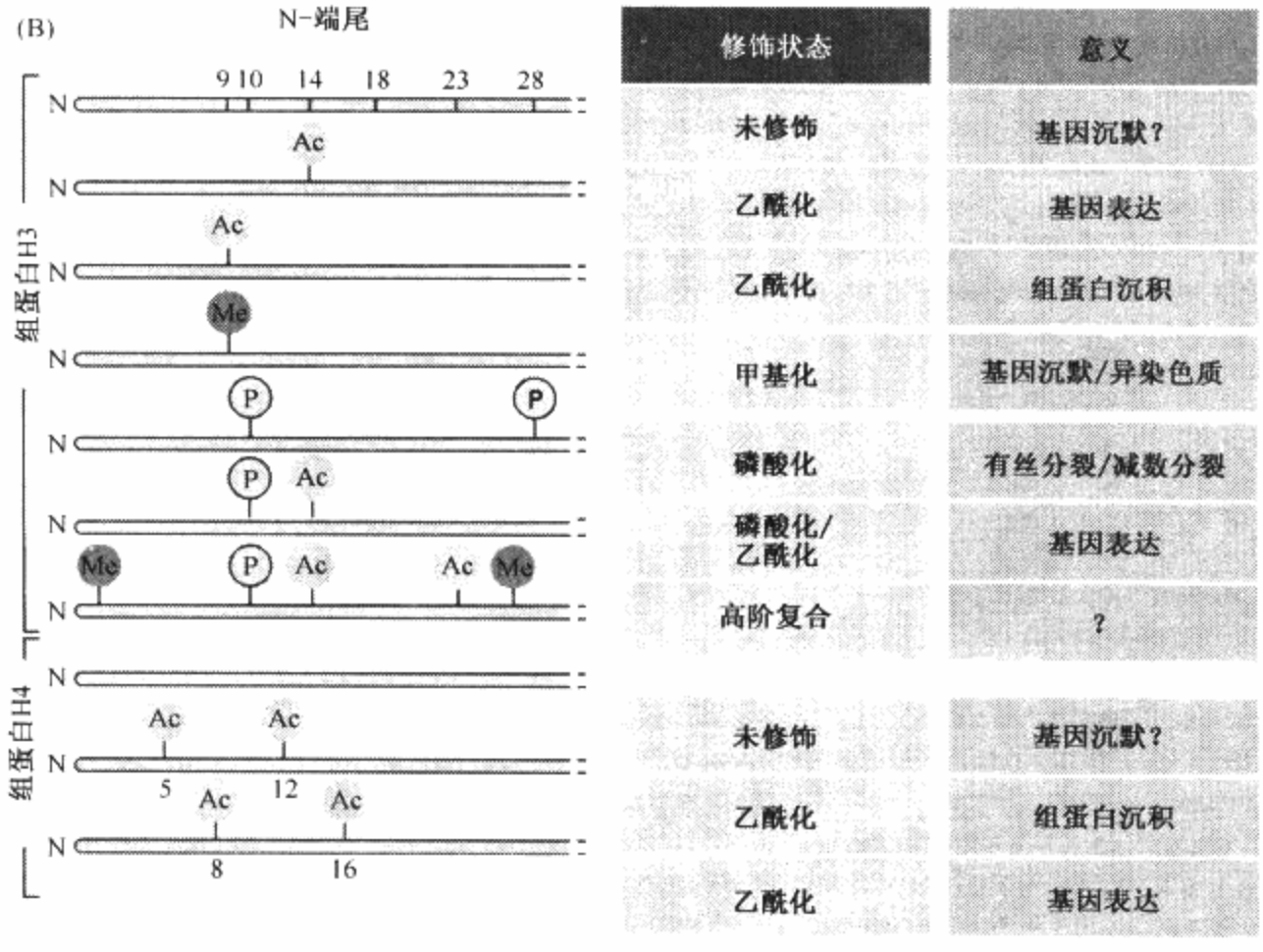


图 4-35　核心组蛋白尾的共价修饰

（A）图中显示了已知的 4 种组蛋白核心的蛋白修饰：Me＝甲基，Ac＝乙酰基，P＝磷酸，u＝泛素。注意有些位置能以不只一种方式（如 H3 上的第 9 位赖氨酸）修饰。大多数这些修饰将组蛋白尾增加了一个比较小的分子；一个例外就是泛素，它是一个含有 76 个氨基酸的蛋白质，并且也用于其他细胞的过程中（图 6-87）。泛素在染色质中的功能还不太清楚；组蛋白 H2B 可以由单个泛素分子修饰；H2A 可由多个泛素的连接加以修饰。（B）组蛋白密码假说。组蛋白可以用不同的修饰组合进行标记。根据这个假说，每一种标记传递给发生标记的染色质片段特殊的含义。只有一些修饰含义是已知的。第 7 章中，我们讨论了基因表达所需的一种蛋白质解读双乙酰化的 H4 尾的方式。在另一个经过详细研究的例子中，一组蛋白识别在第 9 位氨基酸甲基化的一条 H4 尾，这些蛋白质形成一种特别紧密的染色质形式，能够使基因表达发生沉默。组蛋白 H3 的第 14 位赖氨酸和组蛋白 H4 的第 8 和 16 位赖氨酸的乙酰化，通常与基因表达有关，由类型 A 组蛋白乙酰转移酶（HAT）在细胞核内催化进行。相比之下，组蛋白 H4 的第 5 和 12 位赖氨酸和组蛋白 H3 的一个赖氨酸的乙酰化发生在细胞质，在组蛋白已合成以后但未整合到核小体中之前；这些修饰由类型 B 组蛋白乙酰转移酶催化。这些经过修饰的组蛋白在 DNA 合成以后堆积在 DNA 上。（图 5-41），之后不久它们的乙酰基团就由组蛋白脱乙酰酶（HDAC）水解下来。因此这些位点的乙酰化标志着新复制的染色质。根据局部染色质结构的其他特征，一条组蛋白尾上特定位置发生修饰能够代表不同的意义。例如，组蛋白 H3 的第 10 位磷酸化，不仅与有丝分裂和减数分裂中发生的染色体凝聚有关，而且与某些基因表达有关。有些组蛋白尾修饰是相互依赖的。例如，H3 的第 9 位甲基化阻断了 H3 的第 10 位磷酸化，反之亦然

特别恰当的安排形式。

修饰（以及除去修饰）组蛋白尾的酶与染色质重建复合体一样，通常都是多亚基蛋白质，并受到严格的调控。它们由其他线索，尤其是序列特异的 DNA 结合蛋白，引导至染色质特定区域。因此我们可以想像到反复的组蛋白尾修饰和去修饰是如何让染色质结构发生动态变化的——局部压缩或解压缩染色质，此外，还吸引了每种修饰状态特异的蛋白质。有可能组蛋白修饰酶和染色质重建复合体协同作用，压缩和再压缩了染色质。例如，有证据表明特定的组蛋白尾修饰吸引特定类型的重建复合体。此外，有些染色质重建复合体含有组蛋白修饰酶作为亚基，从而直接将两个过程联系起来。

小结

一个基因就是DNA分子中的一段核苷酸序列，是合成蛋白质、结构RNA和催化RNA分子的功能单位。在真核生物中，蛋白质编码基因通常由一串交替排列的内含子和外显子组成。一条含有许多线性排列基因的冗长DNA分子形成了染色体。人类基因组有3.2×10^9个DNA核苷酸对，分属22条不同的常染色体和2条性染色体。只有一小部分染色体DNA编码了蛋白质或者起结构和催化作用的RNA。染色体DNA分子还含有其他三种在功能上很重要的核苷酸序列：复制起点和端点，它们能使DNA分子完整复制；以及中心粒，它能使子代DNA分子附着到有丝分裂纺锤体上，从而确保了在细胞周期的M期，能精确地分离到子代细胞中去。

真核生物内的DNA紧密地结合到等量的组蛋白上，结果形成了重复排列的DNA-蛋白质颗粒，称为核小体。核小体由DNA双螺旋包裹一个组蛋白八聚体核心而来。虽然核小体在DNA上的定位并不规则，但为了形成30nm纤维，通常核小体一起包装成半规则排列（在组蛋白H1分子的帮助下）。尽管染色质高度压缩，它的结构却必须是高度动态变化，以便让细胞能接触到DNA。为了实现这一目的，一般存在两种策略来可逆地改变染色质结构：ATP驱动的染色质重建复合体，以及在4种核心组蛋白N端尾上进行的酶催共价修饰。

染色体的总体结构

讨论过形成30nm染色质纤维的DNA和蛋白质分子以后，现在我们来讨论更高整体水平的染色体组织结构。以30nm纤维存在时，人类染色体仍然长达0.1cm，比细胞核直径宽100倍。显然，必定还存在更高水平的折叠，即便对于有丝分裂间期染色体也同样如此。这种更高水平的包装是染色体结构中最让人感兴趣的一个方面，但也是了解得最少的一个方面。尽管这种包装的分子基础在很大程度上仍然是个谜，但是它几乎肯定涉及将30nm纤维折叠成一系列的环和盘绕，这些我们在下面可以看到。我们有关这种更加有序的包装的讨论开始了染色体组织构造中一个重要的主题：间期染色质的结构是易变的，在任何时候都要裸露出细胞直接需要的DNA序列。

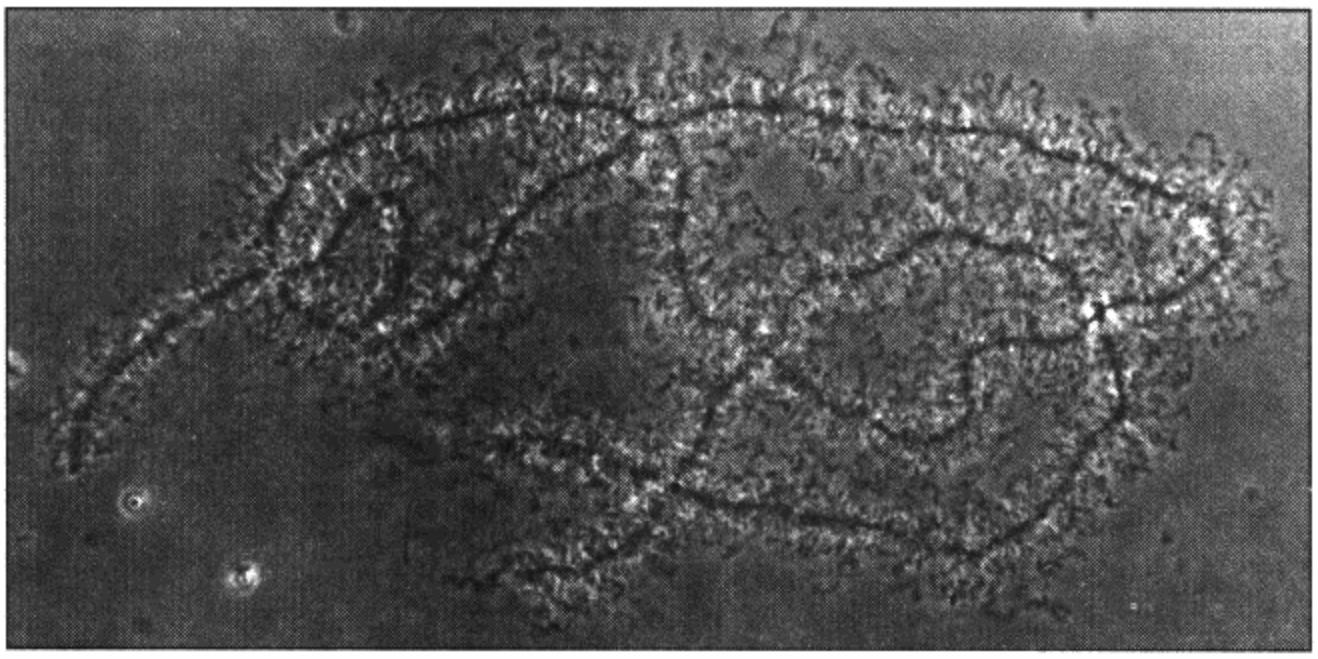
首先，我们描述几个不常见的例子，在这些例子中能容易地观察到间期染色质整体结构和组织形式。我们认为这些罕见的情况可能代表了所有间期染色体结构的特征。接着，我们描述构成典型间期染色体不同形式的染色质。最后，我们讨论在有丝分裂过程中，间期染色体发生的进一步压缩。

灯刷染色体含有解凝聚的染色质环

有丝分裂间期细胞的大部分染色体都太细太乱而不容易清楚地观察。然而在一些少见的情况下，也可以看到**间期染色体**具有精确界定的高级有序结构，据认为这些高级有序结构的某些特征是所有间期染色体的代表。例如，在生长的两栖动物卵母细胞中（未成熟的卵子），有丝分裂期间配对的染色体基因表达非常活跃，它们通常形成僵硬而伸展的染色质环。这些所谓的**灯刷染色体**（已知最大的染色体），即便在光学显微镜下也清晰可见，能够观察到它们巨大的染色质环从一条线性的染色体轴向四周发散出来（图4-36）。

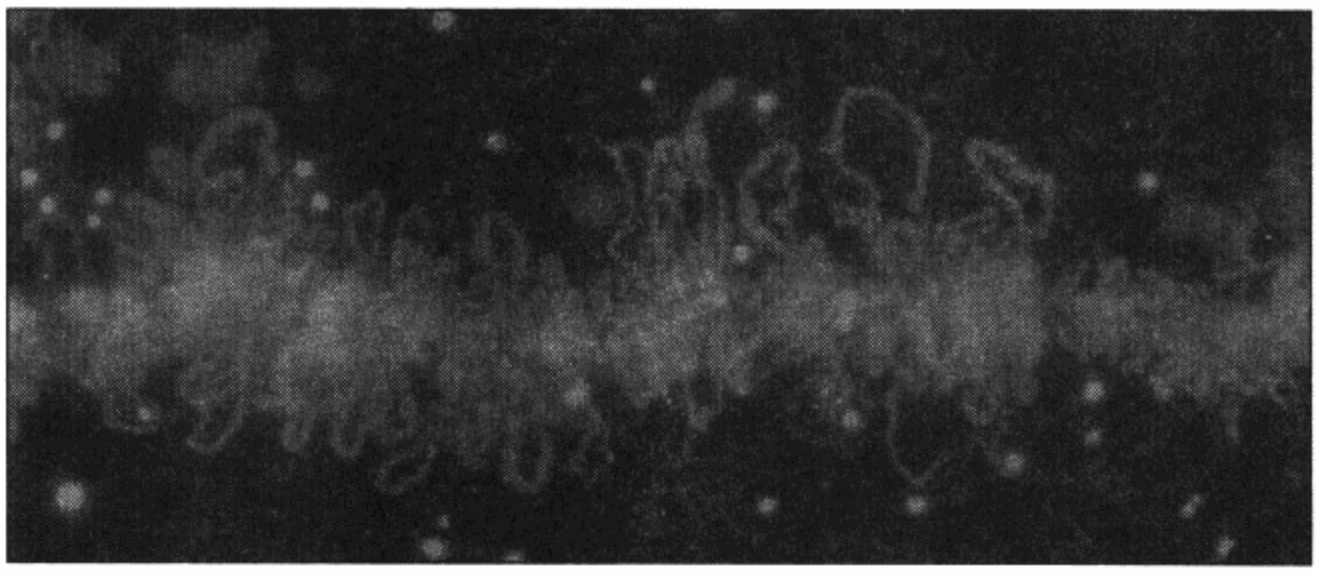
图4-37说明了灯刷染色体的组织情况。一个给定的环总是含有相同的DNA序列，并且当卵母细胞不断生长时，它始终保持同一种方式伸展着。其他实验表明DNA环上存在的大

(A)



0.1 mm

(B)



20 μm

图 4-36　灯刷染色体

（A）两栖类卵母细胞中的一个灯刷染色体的光学显微照片。在卵母细胞分化早期，每个染色体进行复制，为减数分裂作准备，复制出同源染色体配对形成这种高度伸展的结构，它总共含有 4 条复制的 DNA 分子或染色单体。灯刷染色体时期持续数个月或数年，同时卵母细胞持续增加其最终发育成新个体所需的营养物质。（B）显示两栖类灯刷染色体一部分的荧光光学显微照片。利用基因表达某一步骤中加工 RNA 的蛋白质的抗体，将活跃表达的染色体区域染成绿色（讨论见第 6 章）。圆形颗粒认为是对应于 RNA 剪接装置的巨大复合体；这也将在第 6 章中讨论[（A）经 Joseph G. Gall 许可；（B）经 Joseph G. Gall 和 Christine Murphy 许可]

多数基因都在活跃表达[图4-36(B)]。然而大部分 DNA 都不在突环上，而是位于轴上，保持高度凝聚，形成**染色粒**，一般它们不表达。灯刷染色体例证了一个本章重现的主题——如果细胞正在利用染色质某段区域 DNA，这段染色质就呈现伸展的结构；否则，就发生凝聚。在灯刷染色体中，这种调控的结构单位是巨大而且得到精确界定的环。

相对而言，只有极少的物种经过特化产生灯刷染色体。然而，如果将来自正常情况下不形成灯刷染色体的生物的 DNA（如来自某种鱼的 DNA），注射到两栖动物卵细胞内，也会形成灯刷染色体。在这种实验的基础上，有人提出所有真核生物的间期染色体都排列成环，它们在正常情况下太小而且太脆弱，以至于不容易观察到。今后有可能小心地从哺乳动物如小

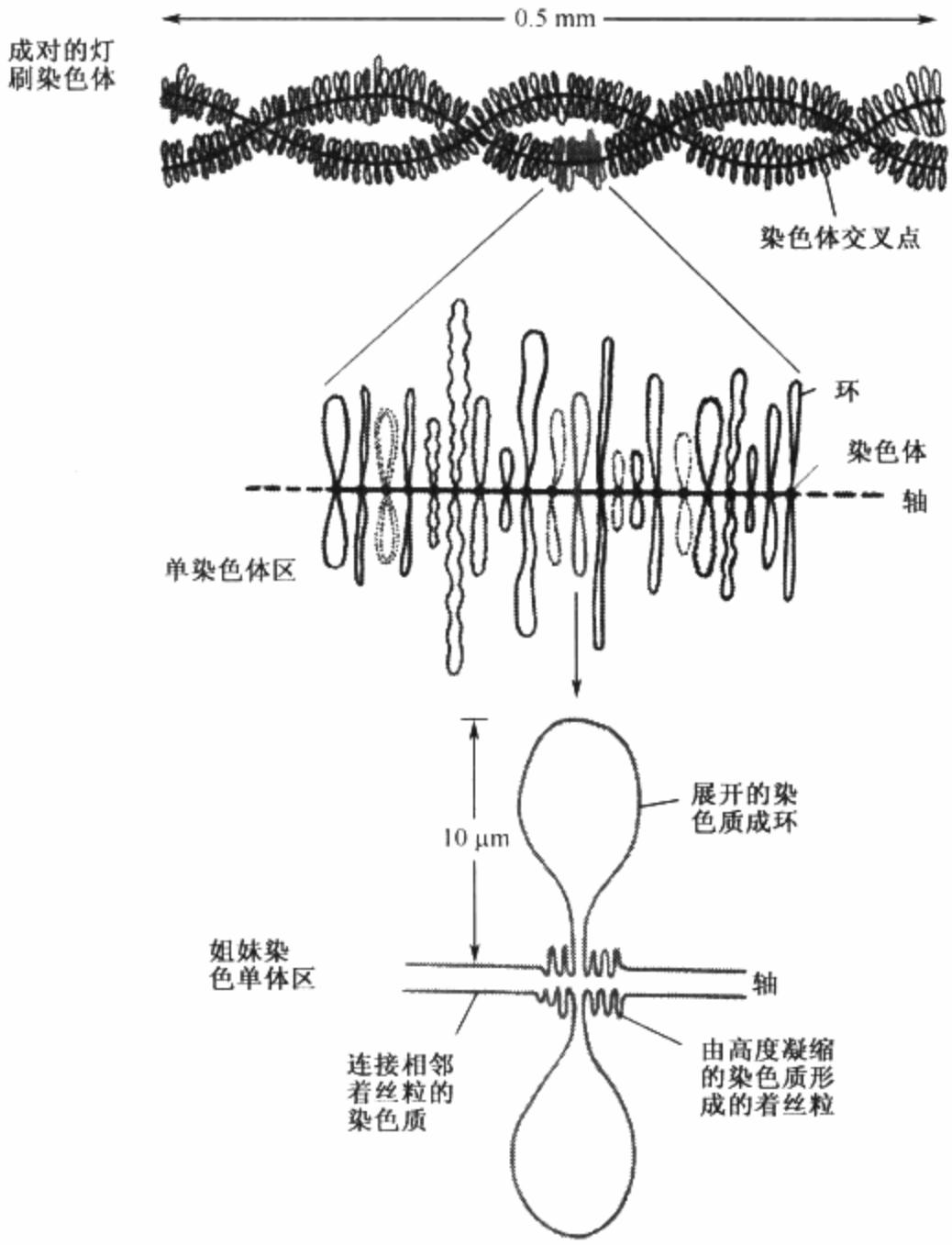


图 4-37　灯刷染色体的结构模型

在许多两栖动物中一套灯刷染色体含有总共约 10 000 个染色质环，尽管每条染色体中的大部分 DNA 都保持高度凝聚成染色粒。每个环对应一段特定的 DNA 序列。由于图的上面所示的两条染色体每个都由两条紧密并列起来的新复制染色体组成，所以每个细胞中每个环存在 4 个拷贝。该四链结构是卵母细胞的减数分裂双线期的特征；见图 20-12

鼠中取出 DNA，将其引入两栖动物卵母细胞后，形成灯刷染色体。这样可以将环结构、基因排列和 DNA 序列具体地联系起来，我们就能着手了解环的包装与 DNA 序列之间的联系。

果蝇多线染色体排列成交替的带和间带

某些昆虫细胞也含有的特化的容易观察到的间期染色体，虽然这种特化与形成灯刷染色体的不同。在某些幼蝇细胞中，有许多细胞经过多轮 DNA 合成，而细胞却不分裂，结果长成一个巨大的细胞。所形成的巨大细胞比全部的正常 DNA 量多几千倍。DNA 含量高于全部正常 DNA 量的细胞称为多倍体，它含有数目增加的标准染色体拷贝。在许多果蝇幼虫的分泌细胞中，所有的同源染色体拷贝并排在一起，就像插入盒内的吸管一样，形成了一条**多线染色体**。有些巨大的昆虫细胞内，多线染色体可以散布存在，形成一个常规多倍体细胞，这个事实表明这两种染色体状态是密切相关的，且多线染色体的基本结构必定与正常染色体的结构相似。

多线染色体通常容易在光学显微镜下观察到，因为它们如此巨大，也因为精确地并行排列的单股染色体之间的附着力大大延长了染色体轴，并防止了缠结的形成。多染色线现象在

图 4-38　果蝇唾腺细胞整套多线染色体

为了便于观察，这些染色体已在显微镜载玻片上压扁后伸展开。果蝇有 4 条染色体，这里所示的是 4 个不同的染色配体对。但是每条染色体与同源物的紧密配对（这样每个染色体配对看起来是单一的结构），在大多数细胞核中并不是这样的（除了在减数分裂间期）。这 4 条多线染色体通常由它们的着丝粒附近区域连接起来，这些区域聚集在一起，形成单个巨大的染色中心（粉红色区域）。但是在此制备过程中，染色中心因挤压过程而断裂成了两半

（摘自 T. S. Painter, J. Hered. 25：467～476，1934）

果蝇幼虫唾腺细胞中研究得最多，它们所含的 4 条果蝇染色体中，每一条的 DNA 经过 10 轮复制，子染色体却不分离，这样 1024（2^{10}）条相同的染色质链就并排起来（图 4-38）。

在光学显微镜下观察多线染色体时，可以看到不同的交替排列的暗带和明亮的间带（图

4-39）。每条带与间带都代表一套排列在染色体中的 1024 个相同的 DNA 序列。多线染色体中大约 95%的 DNA 位于带中，而剩余 5%位于间带中。位于每个带中的染色质呈暗色，可能是因为它比位于间带的染色质压缩程度更高，也可能是因为它含有更高比例的蛋白质，或者两种情况兼而有之（图 4-40）。根据大小不同，每条染色质上单个带据估计含有3000～300 000 个核苷酸对。果蝇多线染色体的带可以通过它们不同的厚度和间隔加以区分，并将它们每一个都赋予一个数字从而得到整条染色体“图”。一套完整的果蝇多线染色体中，大约有 5000 条带和 5000 条间带。

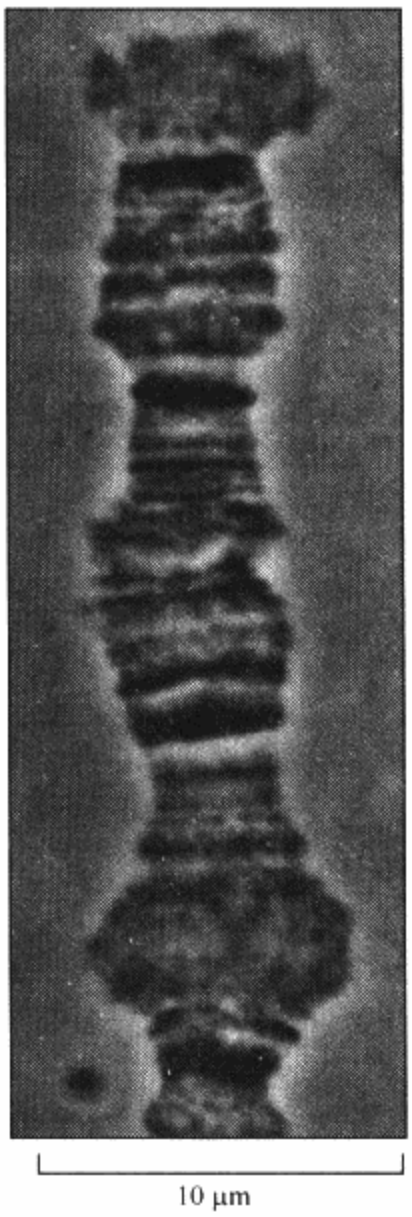


图 4-39　果蝇唾腺多线染色体一部分的光学显微照片
可以清楚地看到带和间带交替产生的不同带型。这些带是在间期染色体中发生染色质凝聚程度增高的区域。虽然它们只能在多线染色体中观察到，但认为它们反映了大多数真核生物中都常见的一种结构（经 Joseph G. Gall. 许可）

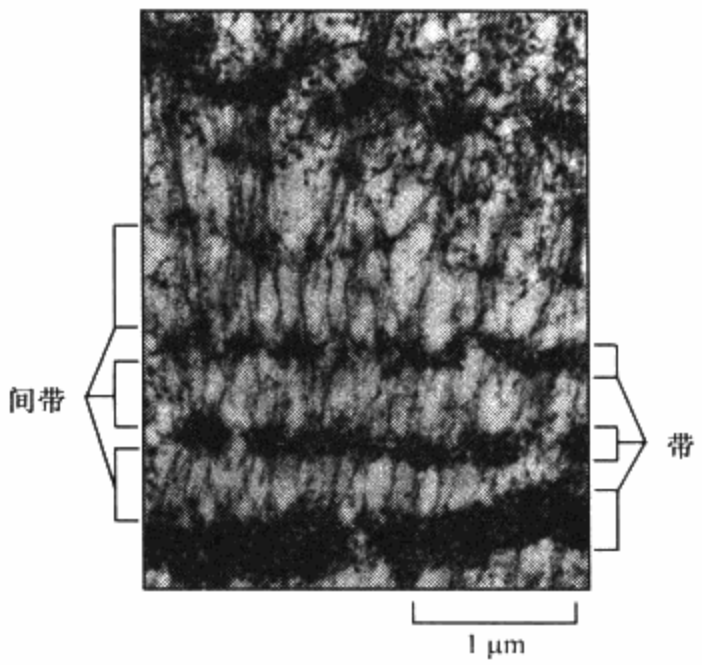


图 4-40　在切片中观察到的果蝇多线染色体一个小截面的电镜显微照片
可以容易地分辨出厚度非常不同的带，它们由间带分隔，含有凝聚程度较小的染色质（经 Veikko Sorsa 许可）

在多线染色体中，带和间带都含有基因

果蝇多线染色体中见到的带与间带的增殖模式意味着这些间期染色体是以高度有序的方式组织起来的。从 20 世纪 30 年代以来，科学家对这种组织的实质一直存在争议，现在我们仍然没有一个明确的答案。由于曾经认为果蝇染色体中带的数量与基因组基因数量是大致相等，所以最初人们认为每个带可能对应单个基因，然而我们现在知道了这种简单的想法是不正确的。果蝇的基因数量比染色体带多将近 3 倍，并且基因在带和间带上都有存在。而且有些带含有多个基因，还有些则看起来完全缺少基因。

看起来这种带-间带模式可能反映了染色体上基因表达和染色质结构的不同水平，压缩程度较小的间带比压缩程度更高的带中的基因表达活跃得多。无论在什么情况下，都认为果蝇多线染色体的显著外形特征，反映了所有间期染色体中发现的染色质压缩的异源性。下一部分我们将看到，当带中的单个和多个基因开始活跃表达时，它的外形怎样发生显著的改变。

单个多线染色体的带能作为一个单位进行解折叠和重折叠

果蝇多线染色体中调控基因表达的一个主要因素就是昆虫类固醇激素——**蜕皮激素**，它的浓度在幼虫发育过程中呈周期性的上升和下降，当蜕皮激素浓度上升时，它们就诱导幼虫每一次蜕皮和化蛹所需的蛋白质编码基因表达。这种生物从一个发育阶段进入到另一个阶段时，由于新基因开始表达，旧基因关闭，所以就会出现不同的**染色体胀泡**，旧的染色体胀泡则消失（图 4-41）。当这种胀泡还比较小，带型仍然能加以分辨时，对它进行观察研究后表明，似乎大部分胀泡是因单个染色体带解凝聚而产生的。

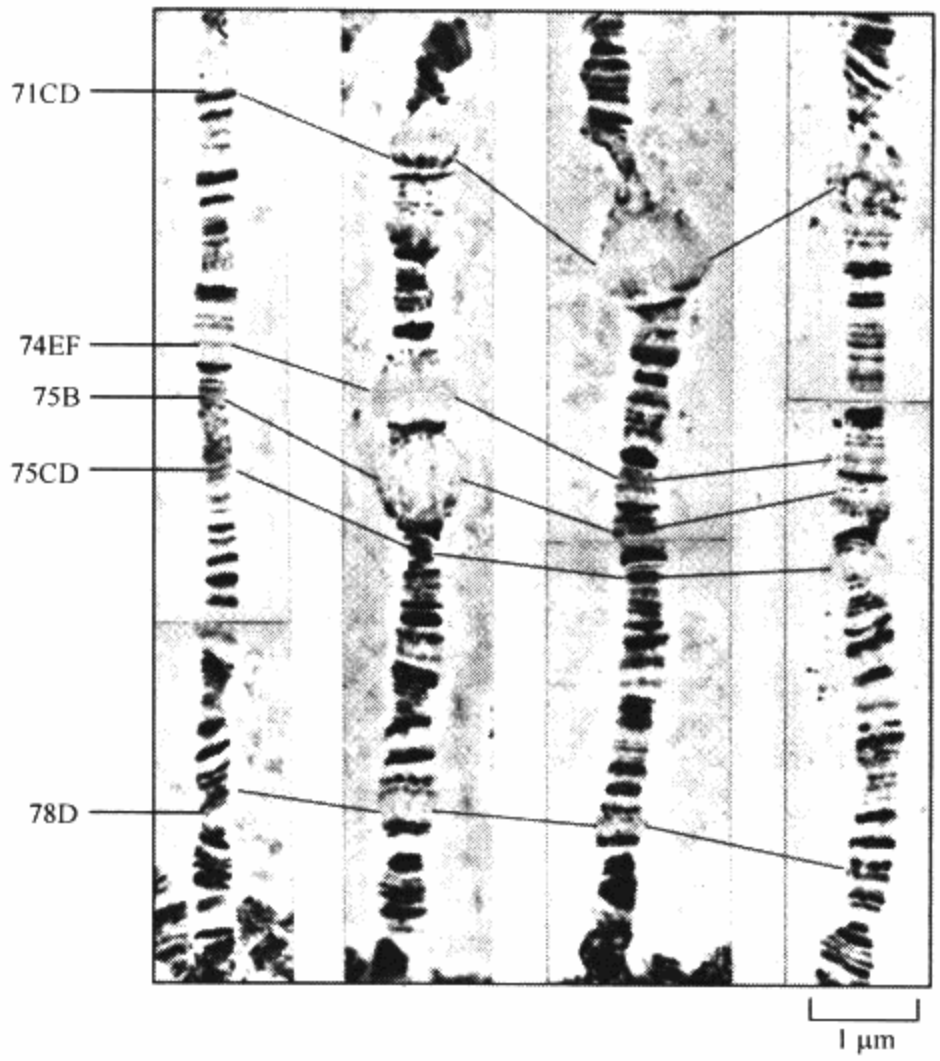


图 4-41　染色体胀泡

这些延时照片显示了在幼虫发育间期果蝇多线染色体中胀泡是如何产生再如何消失的。图中所示为 3 号染色体左臂上的一个区域。它显示了唾腺细胞中 5 个很大的胀泡，每个只在发育的短时期内有活性。照片中显示的一系列变化发生在 22h 内，在组织发育过程中以可再生的模式出现。照片的左边是标示带的命名（经 Michael Ashburner 许可）

组成胀泡的单个染色质纤维可以用电子显微镜观察到。由于技术方面的原因，观察来自另外一种昆虫 *Chironomus tentans*（一种摇蚊）的多线染色体更容易些。某些胀泡，称作摇蚊唾腺多线染色体的巴尔比亚尼环的电子显微照片，显示出排列在其中的染色质（图 4-42 和图 4-43），与前面讨论的在两栖动物灯刷染色体中观察到的非常相似。其他的实验表明，每个环含有一个基因。当基因不表达时，这个 DNA 环呈现出增厚的结构，可能是一个折叠的

30nm 纤维，但是，一旦基因正在表达，这个环就伸展开来，这两种环可能都含有 4 个核心组蛋白以及组蛋白 H1。

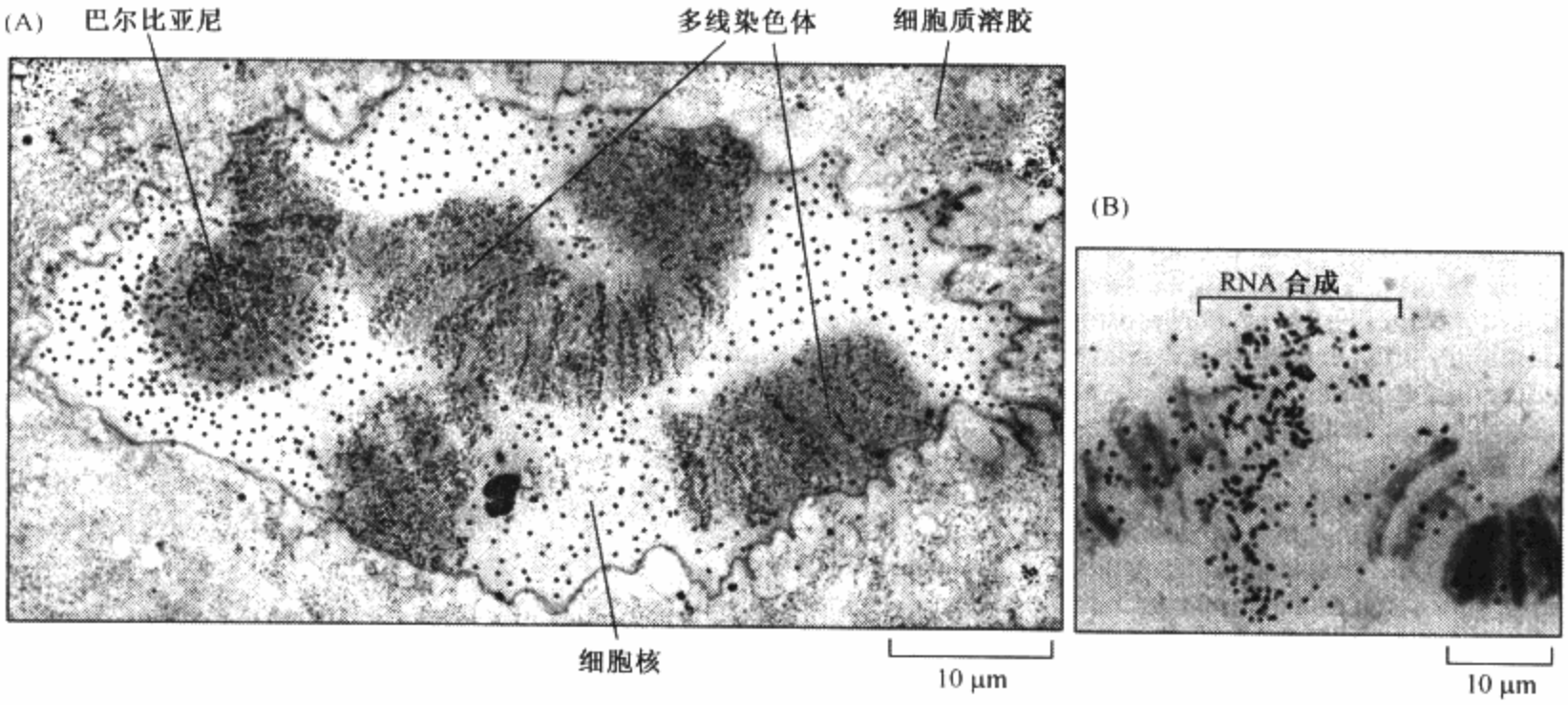


图 4-42　染色体胀泡中的 RNA 合成

(A) 来自一种昆虫，(*C. tetans*)，唾腺的多线染色体。第 1 章已概述并将在第 6 章详细描述，基因表达第一步就是利用 DNA 作为模板合成 RNA 分子。在此电镜照片中，红色表示从一个巴尔比亚尼环基因新合成的 RNA。细胞接触 BrUTP（一种 UTP 类似物）的短时脉冲，这些 BrUTP 就会整合到 RNA 中。然后将细胞固定，利用 BrU 的抗体鉴定出新合成 mRNA。巴尔比亚尼环 RNA 可以通过它们独特的形状与其他 RNA 区分开（图 4-43）。图中蓝点代表加入 BrUTP 之前合成的巴尔比亚尼环 RNA 的位置。整个实验表明巴尔比亚尼环 RNA 在胀泡中合成，然后扩散通过核质。(B) 多线染色体中单个胀泡的放射自显影照片。所示的染色体部分正在进行 RNA 合成，因而标记上 ^{3}H-尿嘧啶[(A)经 B. Daneholt 许可，摘自 O. P Singh et al.,*Exp. Cell Res.* 251：135～146，1999. © Academic Press；(B) 经 Jose Bonner 许可]

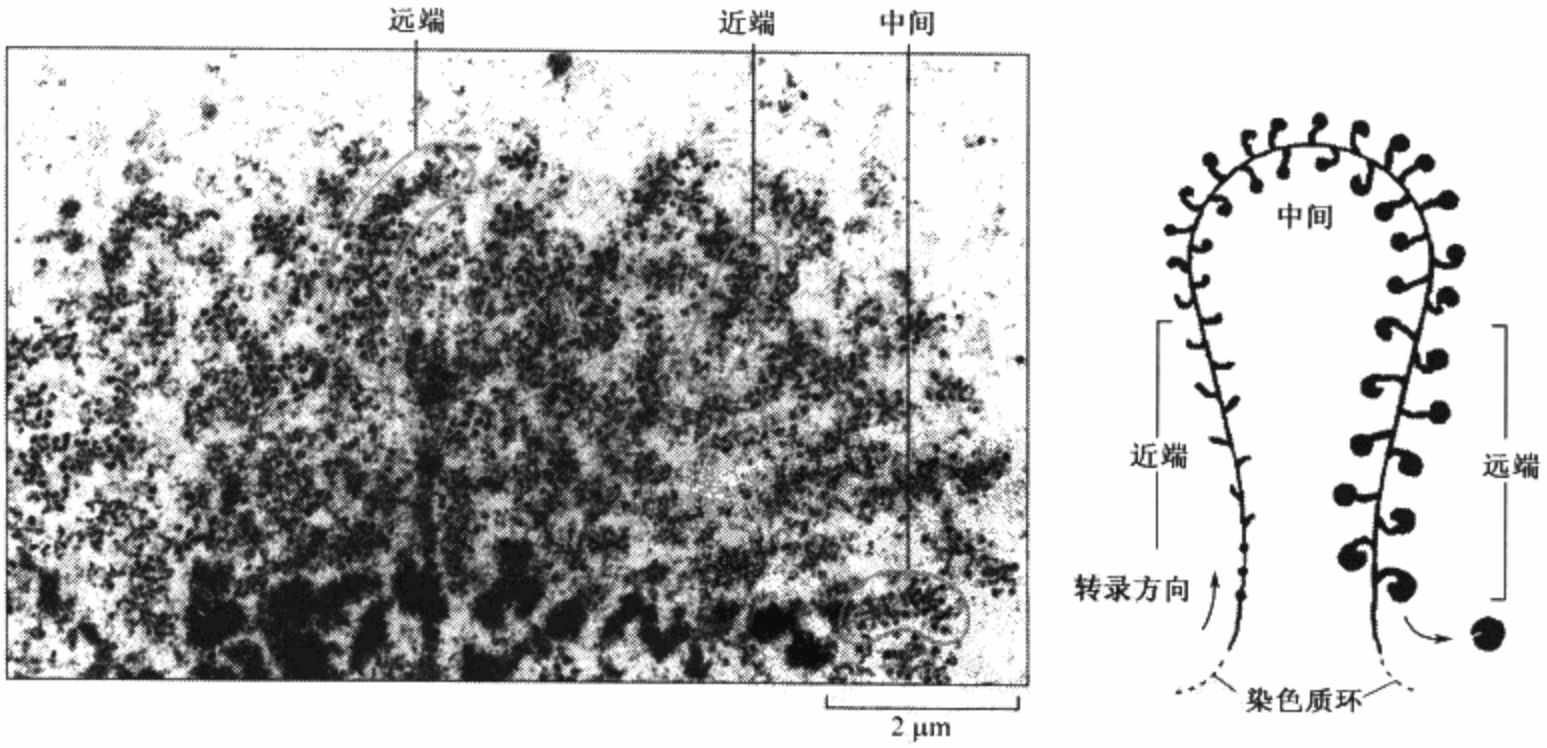


图 4-43　来自 *C. tetans* 的多线染色体

巴尔比亚尼环是基因表达中一种非常活跃的染色体胀泡，这里显示的是该环的染色质薄切片电镜显微照片。巴尔比亚尼环基因编码了分泌蛋白，幼虫用来旋转防护性管道。该染色质排列成环，但是由于样品已切成一部分，所以只能看见这些环的一部分。当 RNA 分子在染色质上合成时，它们与蛋白质分子结合，因而使它们在电镜中呈茎球状。从 RNA-蛋白质复合物大小可以推测 RNA 合成程度（转录）；然后可以从一组切片电镜显微照片（如这里所显示的）重构整个染色质环（右边图示）(经 B. Daneholt 许可，摘自 U. Skoglund et al.,*Cell* 34：847～855，1983，© Elsevier)

看来，默认的环结构可能是一个折叠的 30nm 纤维，而且无论基因在什么时候表达，组蛋白修饰酶、染色质重建复合体和基因表达所需的其他蛋白质，都可能帮助将该纤维转变成更加伸展的形式。在电镜显微照片中，位于解凝聚环两端之一的染色质看起来紧密得多，这与认为单个环组成了染色质结构中独立功能区域的观点是一致的。

尽管有争议，还是有人提出多线染色体中所有 DNA 都排列成环状形式，它们根据所含基因是否表达，而发生凝聚和解凝聚。有可能所有真核生物的所有间期染色体也都包装成一系列有序的环状区域，每个区域含有少量的基因，其表达以协调方式受到调节（图 4-44）。

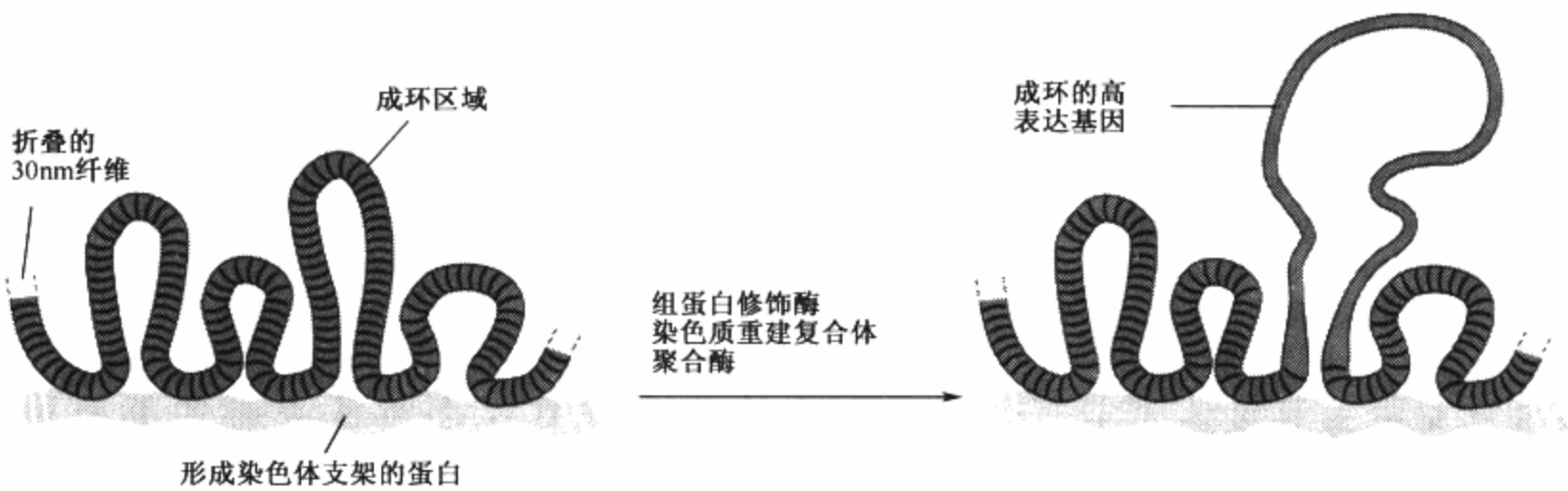


图 4-44　有丝分裂间期染色体结构模型

图中显示的一段间期染色体折叠成一系列环形结构域，每个含有 20 000～100 000 个核苷酸对，它们包含在凝聚成 30nm 纤维的双螺旋 DNA 中。当细胞需要直接利用包装在这些环中的 DNA 时，单个环也能发生解凝聚，可能一部分是通过 30nm 纤维像手风琴一样伸展来实现（图 4-29）。这种解凝聚由直接修饰染色质结构的酶引起，还有直接作用于 DNA 的蛋白质，如 RNA 聚合酶（讨论见第 6 章）。还不清楚折叠的 30nm 纤维是如何锚定到染色体轴上的，但有证据表明染色体环的基础结构成分中富含 DNA 拓扑异构酶，这种酶使 DNA 在锚定时能发生旋转

异源染色质是高度组织的，通常会抑制基因表达

描述了几个少见例子中的间期染色质的一些特征以后，现在我们再来看看在各种生物中都能观察到的间期染色体的特征。20 世纪 30 年代用光学显微镜进行的研究发现许多高等真核生物细胞的间期细胞核中存在两类染色质：一种是高度凝聚的形式，称为**异染色质**；剩余的凝聚程度都较小，称为**常染色质**。常染色质由我们已讨论过的两种染色体结构（30nm 纤维和环区）组成。相比之下，异染色质包括了其他的蛋白质，还可能代表了不为我们所知的压缩程度更高的组织形式。在一个典型的哺乳动物细胞中，基因组将近 10%包装成异染色体。尽管染色体上许多位点存在凝聚的异源染色质，但它们主要集中在特定的区域，包括中心粒和端粒。

大部分折叠成异染色质的 DNA 不含基因。然而，那些包装到异染色质中的基因通常不能表达，这是因为异染色质是异常紧密地压缩起来的。这并不意味着异染色质对细胞无用或有害；下面我们会看到，异染色质所处的区域负责端粒和中心粒（它们缺少基因）的正确功能，它的形成甚至可能有助于保护基因组不被 DNA 的寄生移动元件占据。此外，有些基因要表达就必须位于异染色质区域。事实上，**异染色质**（首次在细胞学中定义）这个词可能包括了许多不同类型的染色质结构，它们共同的特征就是极高程度的组织。由此可见，不能将异染色质看成是不断压缩的死亡 DNA，而应该看作是产生具有不同特征和功能的不同类型的压缩染色质。

异染色质对基因表达的抑制作用，使它即便在那些不能直接观察到的生物体中也容易研究。如果运用实验手段将一个在常染色质中正常表达的基因重新放到一段异源染色质区域中，它将停止表达，称这个基因发生了**沉默**。这些差别是基因表达中的**位置效应**的例子，在此效应中，基因的活性由它在染色体中的位置决定。最先是在果蝇中发现了位置效应，现在许多生物中都发现存在位置效应，人们认为它反映了染色体上染色质结构的不同状态对基因表达的影响。因此，染色体可以看成是不同形式染色质组成的镶嵌体，每条染色质都对细胞赋予其所含 DNA 的能力有特殊影响。

许多位置效应显示出其他的特征，称为**位置效应花斑**，正是它导致了图 4-45 所示的例子中果蝇眼睛出现斑驳，酵母群落发生分区。这些模式的出现，可以是因为原来基因沉默的一群细胞现在重新活化了；一旦活化以后，基因就以这种形式在子代细胞中稳定地遗传。基因也可以在发育早期的常染色质中开始表达，然后经过或多或少的随机筛选以后，包装到异源染色质中，从而使它在细胞及其所有的子细胞中失活。

对位置效应花斑的研究揭示了异染色质两个重要的特征。首先异染色质是动态变化的，它能以可观察到的低频率，扩展到某个区域，然后再恢复原状。其次，染色质所处状态，不论是异染色质还是常染色质，都倾向于遗传给细胞的后代。这两个特征是产生位置效应花斑的原因，如图 4-46 中的解释。异源染色质一旦形成以后，能够自我维持其结构，下一部分，我们将讨论几种解释其实质的模型。

染色体末端存在一种特殊形式的异染色质

与核小体和 30nm 纤维不同，有关异染质结构的了解并不多。它几乎肯定涉及在 30nm 纤维水平上更高程度的折叠，并且需要除组蛋白之外的许多其他蛋白质。简单的酿酒酵母的染色体，尽管在光学显微镜下因太小而看不见，但其异染色质的分子实质可能了解得最清楚。许多用酵母细胞进行的实验表明：从每个染色体末端，染色质向内伸展约 5000 个核苷酸对，它能抑制基因的表达，而且其结构可能对应于更复杂生物的染色体中至少一种类型的异染色质。广泛进行遗传学分析后已经在酵母中鉴定出许多参与这种基因沉默的蛋白质。

酵母的一套沉默**信息调节子**（Sir）蛋白中任何一个发生突变，都能阻止位于端粒附近的基因发生沉默，因而这些基因能够表达。分析这些蛋白质后发现了一个端粒结合 Sir 蛋白复合体，它能识别选定的组蛋白脱乙酰化的 N 端尾[图4-47(A)]。这个复合体中有一种蛋白质是一个高度保守的组蛋白脱乙酰酶，称为 Sir2，它在多种生物包括人类中存在同源物，并可能在形成异染色质特有的组蛋白脱乙酰化模式过程中发挥主要作用。本章前面讨论过，组蛋白尾脱乙酰化作用据认为能促使核小体包装成更紧密的排列，也可能使得它们更难受某些染色质重建复合体作用的影响。此外，异染色质特异的组蛋白尾修饰模式可能吸引其他蛋白质参与形成并维持异染色质结构（图 4-35）。

但是起初 Sir2 蛋白又是怎样分配到染色体末端的呢？一系列其他的实验研究表明存在图 4-47(B)中显示的模型。一种识别酵母端粒特定 DNA 序列的 DNA 结合蛋白也能结合其中一个 Sir 蛋白，导致在端粒 DNA 上组装成完整 Sir 蛋白复合体。接着这个 Sir 蛋白复合体从该位点沿染色体扩展开，从而修饰临近组蛋白的 N 端尾，以形成一个该复合体优先结合的核小体结合位点。可能相邻 Sir 蛋白复合体协同性结合，以及染色质自身回折以便促进 Sir 在邻近区域的结合，共同推动了这种“扩展效应”[图4-47(B)]。此外，异染色质的形成可能需要染色质重建复合体的作用，以便在核小体包装到一起的时候，重新调整它们的位置。

与大多数脱乙酰酶不同，Sir2 需要 NAD^+ 作为辅助因子（图 2-60）。细胞内 NAD^+ 水平

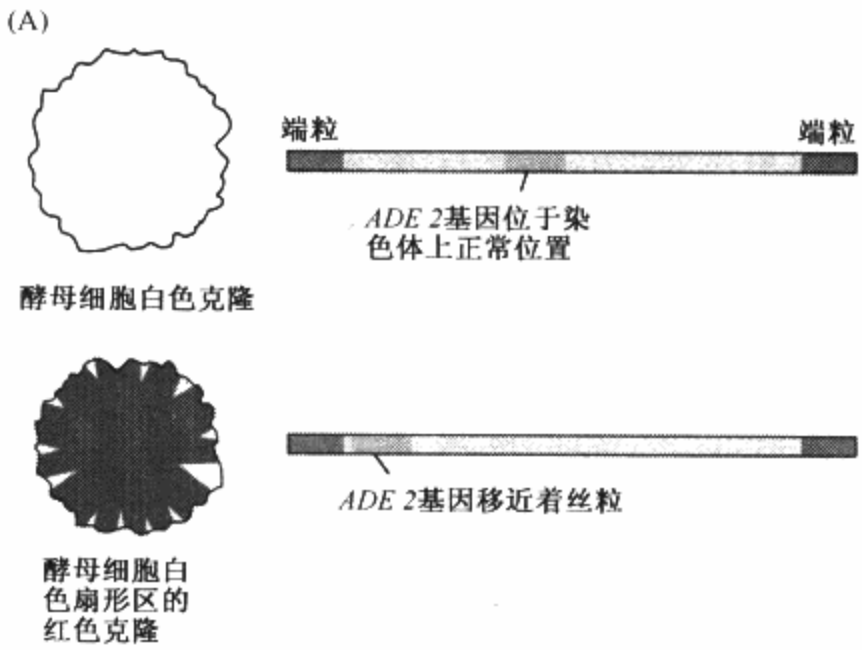


图 4-45　在两种不同真核生物中基因表达的位置效应

（A）处于正常染色体位置的酵母 *ADE 2* 基因在所有细胞中都表达。如果转移到酵母染色体末端附近，据推测在这里会折叠成异染色质形式，那么这个基因在群体中大多数细胞内都不会表达。*ADE 2* 基因编码了腺嘌呤生物合成中的一种酶，*ADE 2* 基因产物缺乏会导致一种红色颜料的累积。因此，表达 *ADE 2* 的细胞群落是白色的，而未表达 *ADE 2* 的细胞组成的群落是红色的。从生长在琼脂表面的红色群落中间，呈扇形扩展的白色扇形区域代表细胞的后代，其中的 *ADE 2* 基因自发地激活了。这些白色扇形区的产生是因为 *ADE 2* 基因附近染色质包装状态变得不再那么紧密，而且这种变化可以遗传。尽管酵母染色体太小，不能在光学显微镜下观察到，但是认为酵母染色体末端的染色质结构与更大的生物染色体中的异染色质具有许多共同的结构特征。（B）位置效应也能在果蝇的 *white* 基因中观察到。*white* 基因调控眼部色素的合成，并根据最先鉴定出该基因的突变命名。*white* 基因正常（$white^+$）的野生型果蝇能正常合成色素，从而眼呈红色，但是如果 *white* 基因发生突变而失活，那么突变体果蝇（$white^-$）就不能合成色素，因此眼呈白色。在果蝇中如果将正常的 $white^+$ 基因转移到异染色质区域附近，眼就呈现斑驳，既有红斑，又有白斑。白斑块代表 $white^+$ 基因由于异染色质影响发生沉默的细胞。相比之下，红斑块代表 $white^+$ 基因表达的细胞，因为在此时，发育早期，异染色质还未形成，它的影响还未扩展到这个基因。与酵母中的情况类似，大群红色和白色细胞的出现表明，该基因的转录活性状态一旦在早期胚胎由染色质包装决定以后，就可以遗传

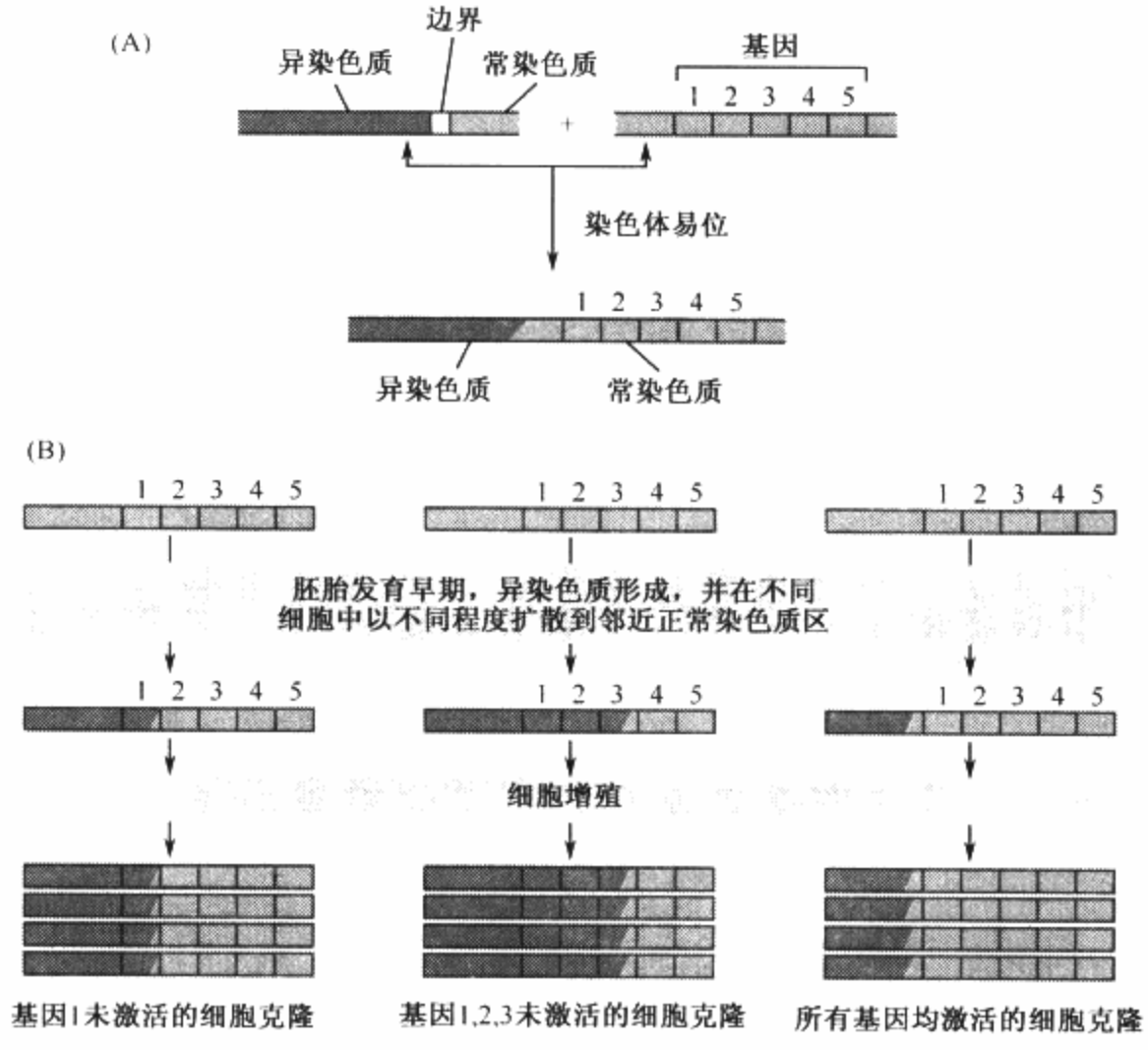


图 4-46　果蝇中位置效应花斑的成因

（A）第 7 章将会讨论，通常特殊的边界 DNA 序列阻止了异染色质（蓝色）扩展到相邻常染色质区域（绿色）。然而，在继承了某些染色体重排结果的果蝇中，就不再存在这个障碍。（B）这种果蝇早期发育过程中，异染色质会扩展到相邻染色体 DNA，在不同的细胞中前进到不同的位置。不久以后这种扩展会停止，但已建立的异染色质模式却遗传下去，这样就产生了大量的后代细胞克隆，它们拥有同样的凝聚成异染色质的相邻基因，因而失活了［因此有些果蝇中出现斑驳的外形；如图 4-45（B）所示］。虽然“扩展”用于描述原先就存在的异染色质周围形成新的异染色质，但这个词并不完全准确。有证据表明在扩展期间，异染色质能“跳过”某些区域，使位于其中的基因免受抑制影响。一种可能性是异染色质扩展能越过某些 DNA 环的碱基，因而绕过了包含在环中的染色质

随细胞营养健康状况发生波动，当细胞缺乏营养时，NAD^+ 水平就会升高。这个特征可能引起端粒异染色质扩展，以应对饥饿（可能使那些并非细胞存活必需的基因表达发生沉默），当条件改善以后，再恢复原状。

上述酵母异染色质的特性可能与更复杂生物的异染色质特性类似。当然，酵母异染色质从端粒的扩展，在理论上与动物中引起位置效应花斑的异染色质运动相似（图 4-46）。而且，这些特性可以用来解释异染色质的遗传力，见图 4-48 中的简图。无论异染色质形成的确切机制如何，显然核小体核心组蛋白的共价修饰在此过程中发挥了关键作用。**组蛋白甲基转移酶**对于许多生物特别重要，它们催化组蛋白上特定的赖氨酸发生甲基化，包括组蛋白 H3 上的第 9 位赖氨酸（图 4-35）。特异结合组蛋白 H3 的这种修饰形式，诱导组装成异染色质的异染色质组分（包括果蝇中的 HP1），能“阅读”这种修饰。有可能细胞利用了一系列不同的组蛋白修饰来区分异染色质和常染色质。

将染色体末端包装成异染色质为细胞提供了许多好处：它有助于保护染色质末端，防止

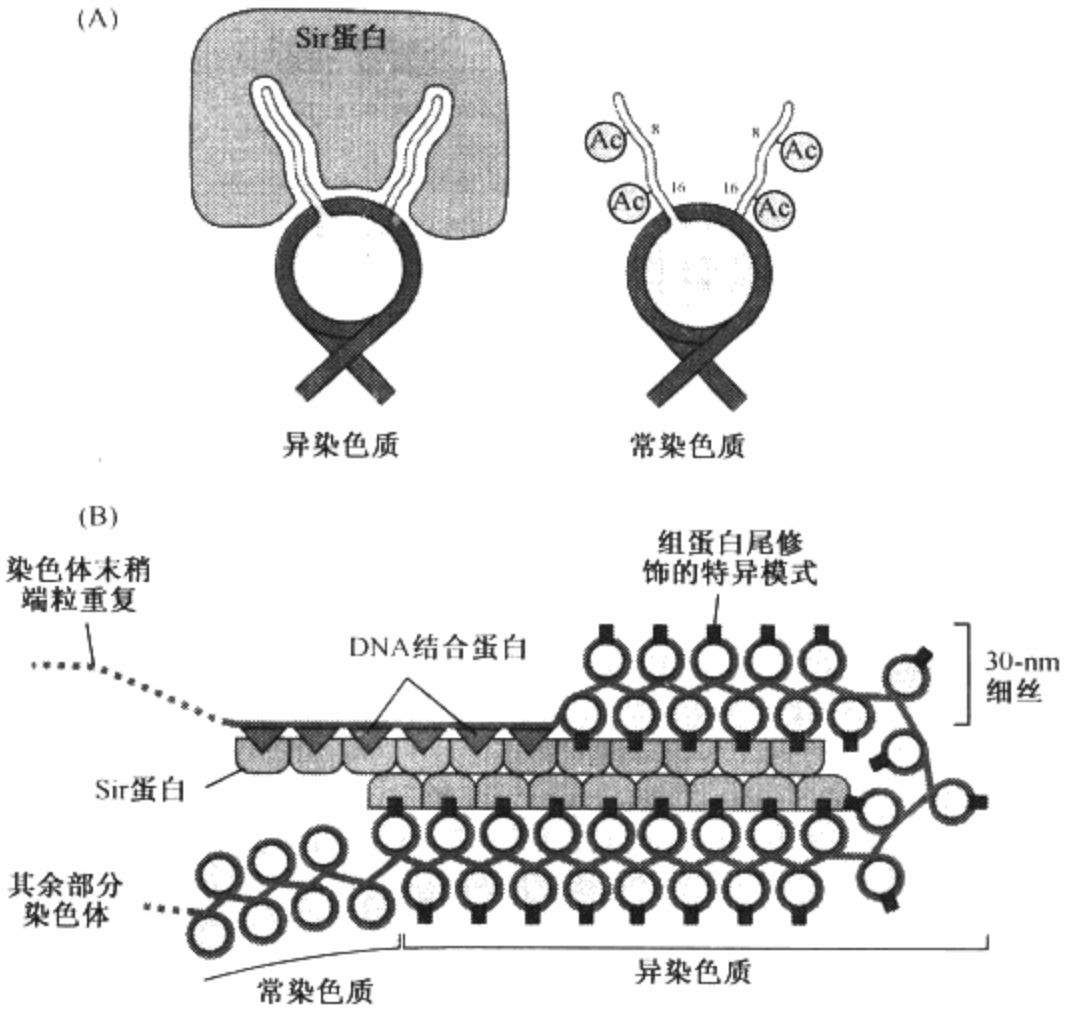


图 4-47　酵母染色体末端异染色质的猜测模型

（A）异染色质一般都乙酰化不足，据推测，组蛋白 H4 乙酰化不足的尾与 Sir 蛋白复合物相互作用，因此稳定了这些蛋白质与核小体的结合。尽管图中所示好像全部都未乙酰化，但还不清楚与 Sir 复合体结合所需的组蛋白 H4 尾部修饰的确切模式。在有些生物中，组蛋白 H3 第 9 位赖氨酸的甲基化也是形成异染色质的关键信号。在常染色质中，组蛋白尾通常高度乙酰化。图中所示 H4 的组蛋白尾只部分乙酰化，但事实上，在常染色质不同的地方乙酰化状态不一样。（B）特殊的 DNA 结合蛋白（蓝色三角形）识别染色体末端附近的 DNA 序列，并吸引 Sir 蛋白，这些蛋白质中有一个（Sir2）是 NAD^+ 依赖的组蛋白脱乙酰酶。这种吸引然后引起 Sir 蛋白复合体沿染色质协同性扩展。随着该复合体的扩展，Sir2 催化的脱乙酰作用在核小体上产生新的结合位点，供更多的 Sir 蛋白复合体结合。还可能形成图中所示的“回折”结构

细胞修复装置将其当作断裂的染色体；它可能有助于调节端粒的长度，并且在有丝分裂期间，它还可能有助于染色体进行精确的配对和分离。

着丝粒也包装成异染色质

异染色质也可以在着丝粒附近观察到，它是在每次细胞分裂中，指导每条染色体移动进入子细胞 DNA 序列（图 4-22）。在许多复杂的生物中，每个着丝粒似乎嵌入一段很长的异染色质中，这段异染色质在整个间期都存在，尽管着丝粒只在有丝分裂期间引导 DNA 运动。有关这种所谓的**着丝粒异染色质**的结构和生物化学特性，知道的还不多，但是与其他形式的异染色质一样，用实验方法将基因放入这段染色体后，它能使其表达发生沉默。除组蛋白（在异染色质中通常乙酰化不足且甲基化）以外，它还含有许多其他的结构蛋白，将核小体压缩成特别凝缩的结构。

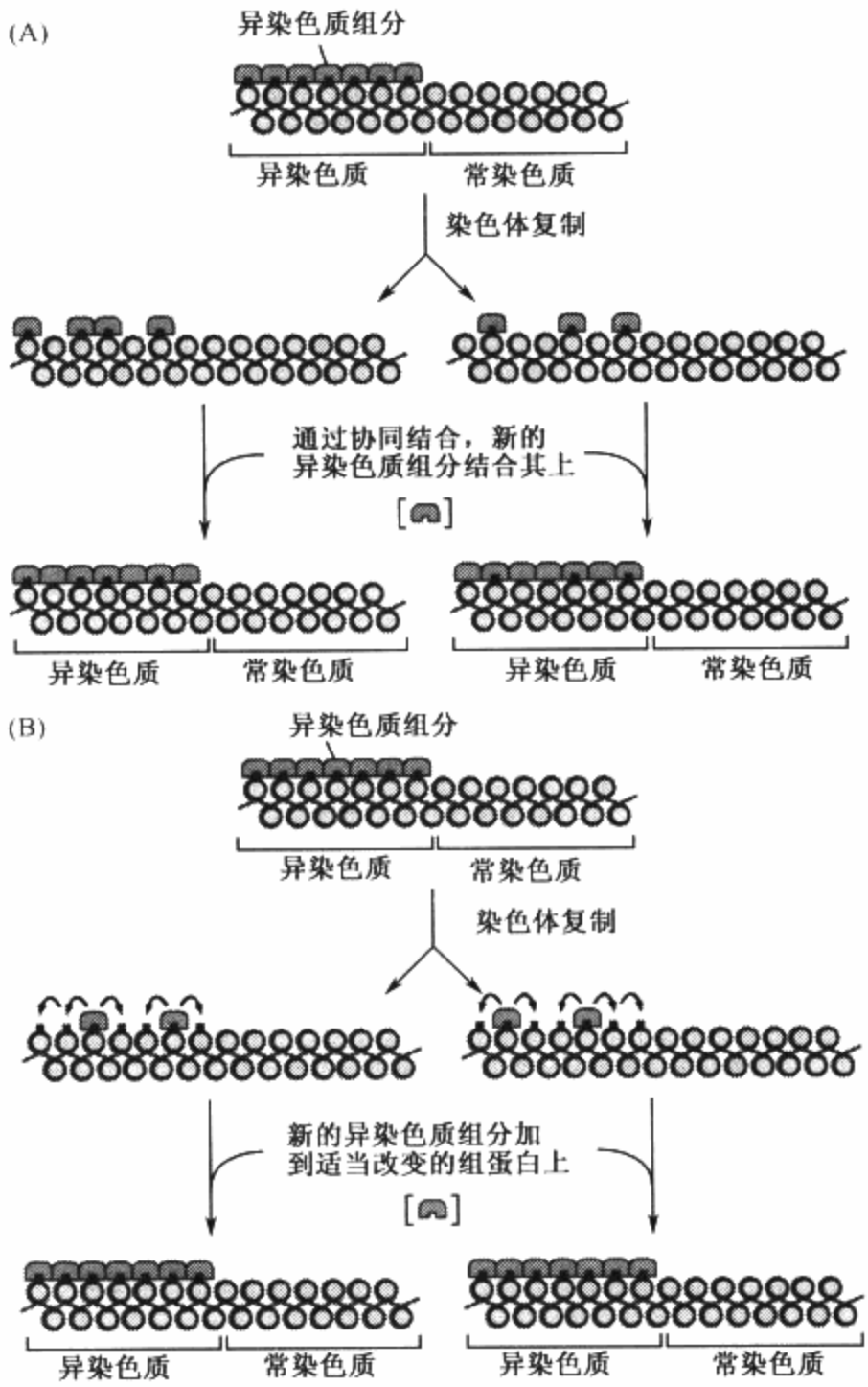


图 4-48　异染色质中紧密包装的 DNA 如何在染色体复制中遗传的两个猜测模型

在两种情况下，特殊的异染色质组分在 DNA 加倍以后都有一半分配到每个子染色体。（A）该模型中，新的异染色质组分协同性地结合到遗传的组分上，从而开始了新异染色质形成过程。其他蛋白质装配上来并且组蛋白发生最终的共价修饰以后（未示出），这个过程就完成了。（B）该模型中，遗传的异染色质组分改变附近新合成的子代核小体上组蛋白修饰模式，产生了新的结合游离异染色质组分的位点，这些组分装配起来形成完整的结构。两种模型都解释了异染色质扩展效应，实际上这两个过程在细胞中都可以自发进行

与端粒一样，我们有关着丝粒的染色质结构最清晰的了解，来源于对简单得多的酿酒酵母着丝粒的研究。在本章前面，我们看到一段大约 125 个核苷酸对的简单 DNA 序列就足以作为着丝粒在这种生物中发挥作用。尽管这段序列不长，却有至少 12 种不同的蛋白质在它上面组装；这些蛋白质包括组蛋白 H3 突变体，它们据认为与其他核心组蛋白一起形成了着丝粒特异的核小体[图4-49(A)]。我们还不清楚这种核小体变体类型赋予细胞什么特性，但是所有的真核生物着丝粒中似乎都存在类似的特殊核小体[图4-49(B)]。酵母着丝粒上的其他蛋白质将其附着到纺锤体微管上，并产生信号确保在有丝分裂继续进行之前这种附着已完成（讨论见第 17 和 18 章）。

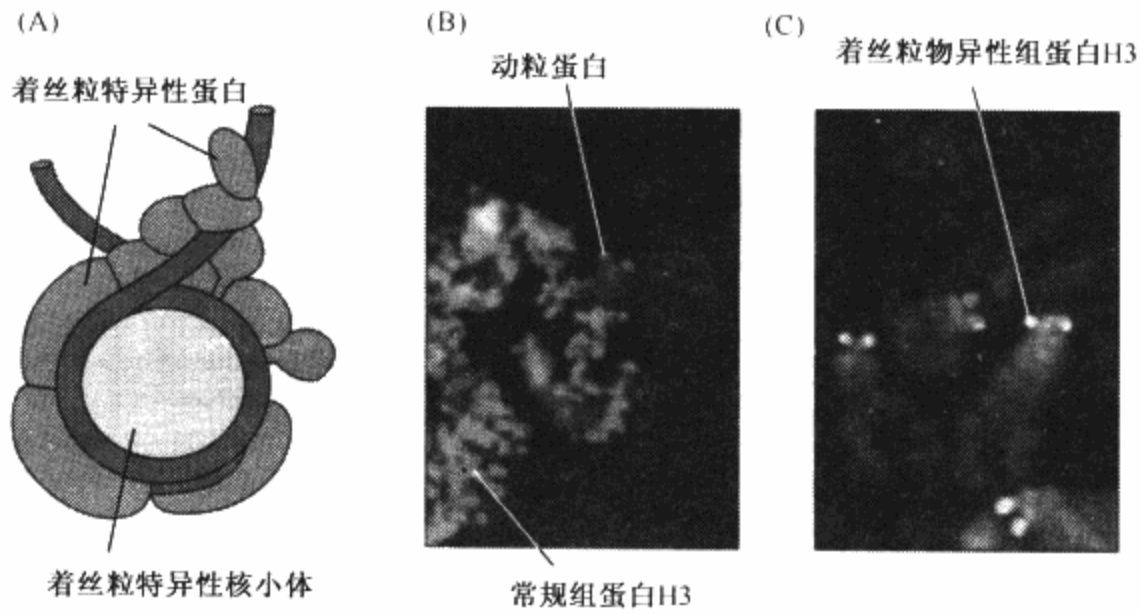


图 4-49　着丝粒上形成的特殊核小体

（A）在酵母着丝粒上装配的蛋白质模型。这个特殊的核小体含有一个 H3 变体（在大多数生物中称为 CENP-A）以及核心组蛋白 H2A、H2B 和 H4。DNA 折叠成核小体促进了其他着丝粒结合蛋白的装配，这些蛋白质形成动粒，用于将着丝粒附着到有丝分裂纺锤体上。（B）果蝇有丝分裂染色体上常规组蛋白 H3 的定位。将常规组蛋白 H4 与荧光蛋白融合，显示出绿色。动粒的一个组分已用一种动粒蛋白特异的抗体染成红色。（C）相同的实验，但是着丝粒特异的组蛋白 H3（而不是常规的 H3）标记成绿色。当红色和绿色同时存在时，染色就变为黄色[（A）摘自 P. B. Meluh et al.,*Cell* 94：607～613，1998；（B）和（C）摘自 S. Henikoff et al.，*Proc. Natl. Acad. Sci.* USA 97：716～721，2000. © National Academy of Science]

更复杂生物的着丝粒比芽殖酵母的着丝粒要大得多。例如，果蝇和人类着丝粒长达数十万个核苷酸对，并且似乎不含着丝粒特异的 DNA 序列。取而代之，大多数主要由重复的短 DNA 序列组成，在人类中它们称为 **α 卫星 DNA**（图 4-50）。但是同样的重复序列也能在染色体其他（非着丝粒的）位置找到，至于它们是如何规定着丝粒区域的，所知甚少。可能先以某种方式播下动粒形成的“种子”，然后形成动粒的全部特殊蛋白质协同性组装起来。似乎在复杂的生物中界定着丝粒的，与其说是由特殊 DNA 序列，还不如说是组装在一起的蛋白质。

着丝粒的形成和维持与异染色质其他区域的形成和维持之间存在许多显著的相似性。整个着丝粒作为一个全有或全无的整体形成，表明播种事件以后，蛋白质高度协同地加入组装。而且一旦形成，其结构作为每轮染色体复制的一部分，似乎直接在 DNA 上遗传。因此，比如说，我们染色体的某些区域含有无功能的 α 卫星 DNA 序列，它们与着丝粒上的序列看起来似乎相同；据推测产生这些序列是因为发生了染色质连接，最初该事件中形成一条具有两个着丝粒的染色体［一种不稳定的双着丝粒染色体，图4-51(A)］。此外，在某些不寻常的情况下，观察到染色体片段上自发形成了新的人类着丝粒（称为新着丝粒）。这些新位点中有些最初都是常染色质，并完全缺少 α 卫星 DNA[图4-51(B)]。

为了解释观察到的这些现象，有人提出，从头着丝粒形成要求最初发生标记（或播种）事件，可能是形成一个特殊的 DNA-蛋白质结构。在人类中，比起其他序列，该事件更容易在 α 卫星 DNA 序列上发生。在染色体分离时，这个标记将进行复制，这样相同的着丝粒在下一轮细胞分裂时，继续发挥作用。在极不常见的情况下，标记在染色体复制后会发生丢失，这时想要再重建着丝粒就非常困难[图4-51(C)]。尽管对有关着丝粒自我更新的性质并没有详细的了解，但是图 4-48 中描述的这种异染色质遗传模型在此同样是关键的。

着丝粒可变性的可能提供了重要的进化优势。我们已经看到染色体进化一部分是通过断

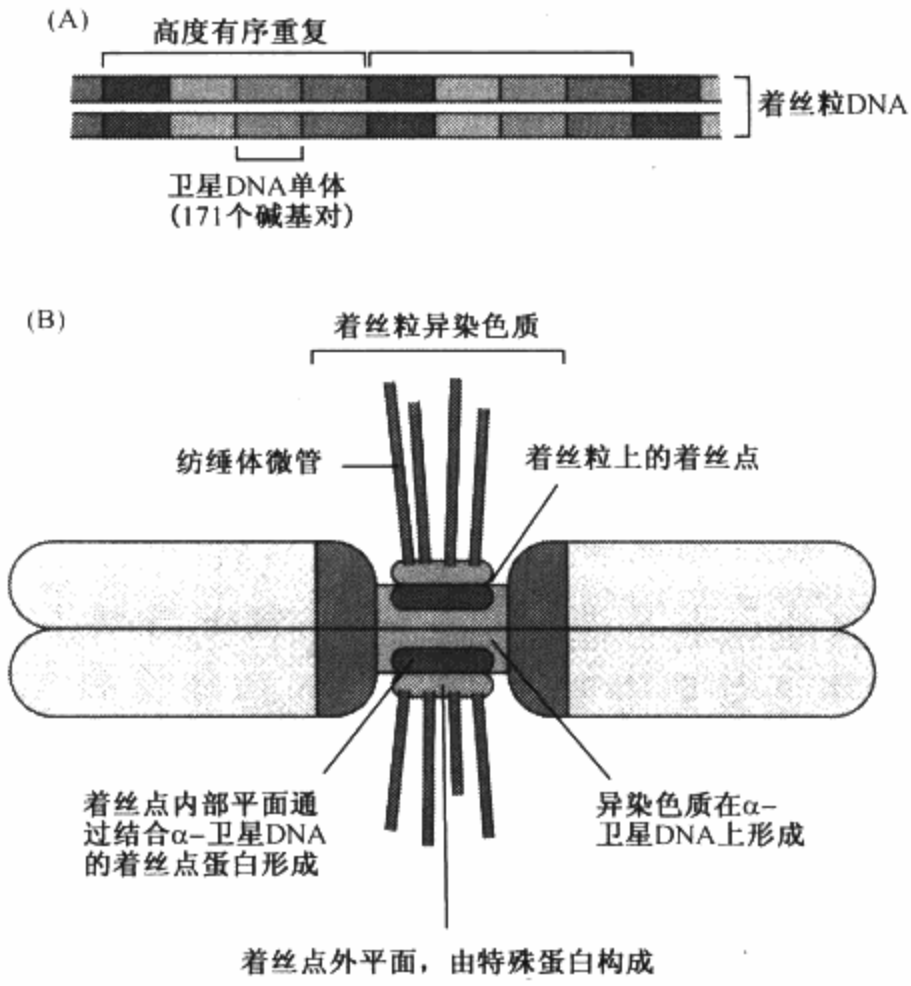


图 4-50　人类着丝粒的结构

（A）α 卫星 DNA 序列的组织，它们在着丝粒上重复达上万次。（B）完整的着丝粒。卫星 DNA 序列（红色）富含 AT，并由大量的序列彼此略有不同的重复组成。蓝色代表着丝粒异染色质侧翼的位置。它含有不同类型的重复组成的 DNA 序列。如图所示，动粒由内侧和外测的圆盘组成，它们由一组动粒蛋白形成。在细胞周期的 M 期，纺锤体微管附着到动粒上（图 4-22）[(B)摘自 T. D. Murphy and G. H. Karpen, *Cell* 93: 317～320, 1998]

裂和再连接进行的（图 4-49）。这些事件中有许多产生具有两个着丝粒的染色体或者根本没有着丝粒的染色体片段。尽管很罕见，着丝粒失活以及它们被从头激活的能力，可能偶尔使新形成的染色体得以保持稳定，因而促进了染色体的进化过程。

异染色质可能提供了一种抵制移动 DNA 元件的机制

包装成异染色质的 DNA 常常由大量的短串联重复序列组成，它们并不编码蛋白质，这一点我们已从上面哺乳动物中心粒的异染色质看到。相比之下，常染色质 DNA 则富含基因和其他单拷贝序列。尽管这种相关性并不是绝对的（有些重复序列排列存在于常染色质，也有些基因存在于异染色质中），但这种趋势表明，有些类型的重复 DNA 可能是形成异染色质的信号。人为地将几百个串联的基因拷贝引入果蝇和小鼠生殖细胞系的实验，支持这一观点。在这两种生物内，这些基因排列通常都沉默了，并且在有些情况下，在显微镜下可以观察到它们形成了异染色质区域。相比之下，如果将相同基因的单拷贝引入染色体同一位置，它们则活跃地进行表达。

这个特征称作**重复诱导基因沉默**，可能是细胞具备的一种机制，用来防止它们的基因组被移动遗传元件占据。这些元件在第 5 章讨论，它们可以增殖，并将自身插入到基因组的各个位置。根据这个观点，一旦形成一串这样的移动元件，含有它们的 DNA 将包装成异染色质，阻止它们的进一步增殖。相同的机制也可能负责形成很长的含有大量的串联重复简单序

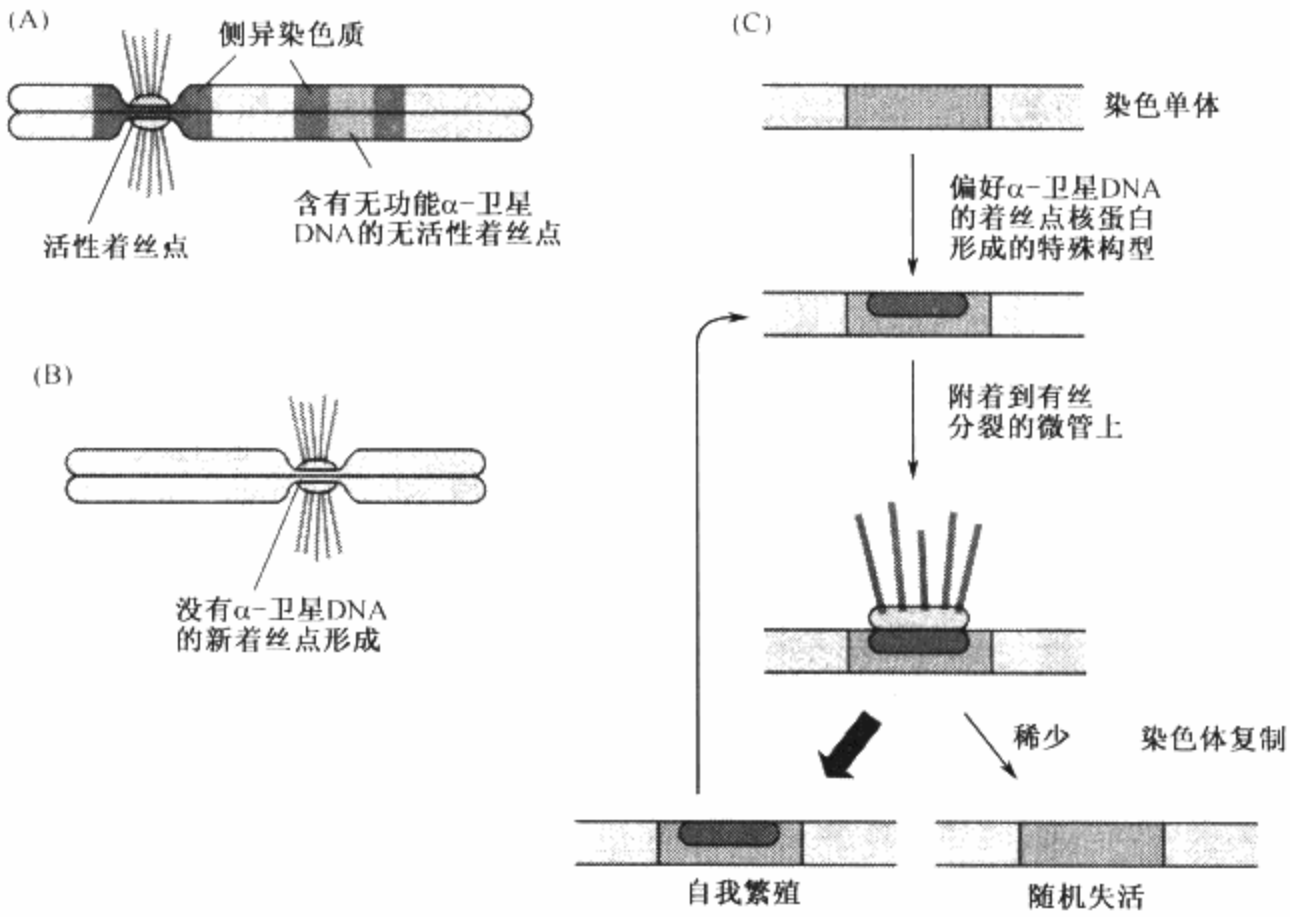


图 4-51　人类的着丝粒形成的柔性

（A）由于古代染色体发生过断裂和再连接，有些人的染色体含有 2 个小卫星 DNA 区域（红色），其中每个可能在最初的染色体中作为着丝粒起作用。通常，这些双着丝粒染色体的增殖不稳定，因为它们没有正确地附着到纺锤体上并在有丝分裂期间分离。在那些存活下来的染色体中其中 1 个着丝粒自发失活，即使它含有全部必需的序列。这样染色体就能够稳定地增殖。（B）一小部分人（1/2000）出生时，在后代细胞中可观察到额外的染色体。这些额外的染色体有些是因为发生断裂而形成，它们完全缺少 α 卫星 DNA，但是在起初是常染色体 DNA 的位置出现了新的着丝粒。（C）解释着丝粒柔性和遗传的模型

图 4-52　处于细胞分裂中期的典型有丝分裂染色体

每条姐妹染色单体含有两条相同的子 DNA 分子之一，它们通过 DNA 复制在细胞周期早期合成

列的异染色质区域，就像在着丝粒周围发生的那样。

有丝分裂染色体由处于最凝聚状态的染色质形成

讨论过间期染色体动态结构以后，我们现在开始讨论 DNA 包装的最后一级水平，也就是观察到的有丝分裂染色体组织结构。除少数几个特殊情况例外（如前面讨论的灯刷染色体和多线染色体），大部分间期染色体结构太过于伸展或缠结成团，以至于不能清楚地观察到它们的结构。相比之下，几乎所有真核细胞的染色体在有丝分裂间期都容易被观察到，因为这时候它们缠绕起来，形成高度凝聚的结构。很明显这种进一步的压缩，虽然使典型的间期染色体长度**仅减少约 10 倍**，却在染色体外观上产生了如此明显的变化。

图 4-52 描绘了处于有丝分裂中期的典型**有丝分裂染色体**。在细胞分裂周期的间期，DNA 复制产生的两条子 DNA 分子单独折叠，形成两条姐妹染色体或**姐妹染色单**

体，它们由着丝粒维系在一起（图4-21）。这些染色体上通常覆盖有不同的分子，包括大量的RNA-蛋白质复合物。一旦剥离这层覆盖物，就可以在电子显微镜下观察到，每条染色单体组装成染色体环，从中心骨架向四周发散出去（图 4-53 和图 4-54）。许多实验证明，在有丝分裂染色体上可见特征的排列顺序至少大致反映了基因在 DNA 分子上的排列顺序。因而可以认为有丝分裂染色体凝聚是染色体包装等级中的最后一级水平(图 4-55)。

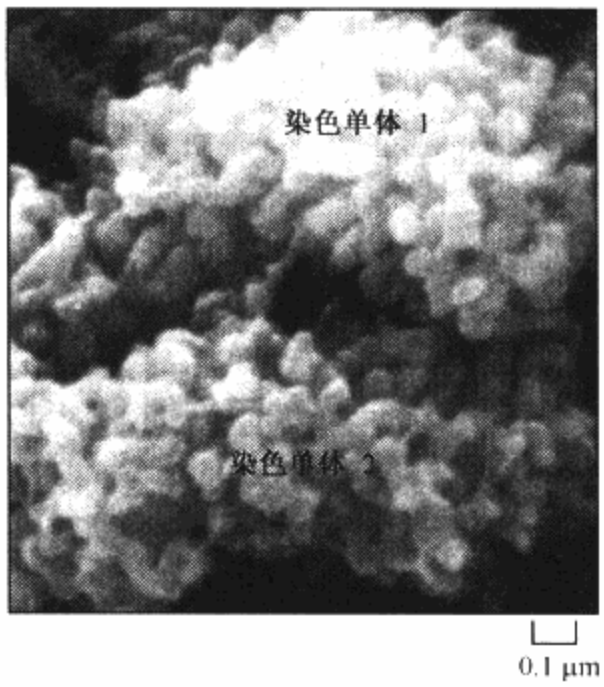


图 4-53　典型有丝分裂染色体一个末端附近区域的扫描电镜照片

每个球形把手样的凸出物代表单独的成环区域的顶端。注意，可以清楚地分辨图 4-52 中所描绘的两条相同的配对染色单体(摘自 M. P. Marsden and U. K. Laemmli, *Cell* 17:849～858, 1979. © Elsevier)

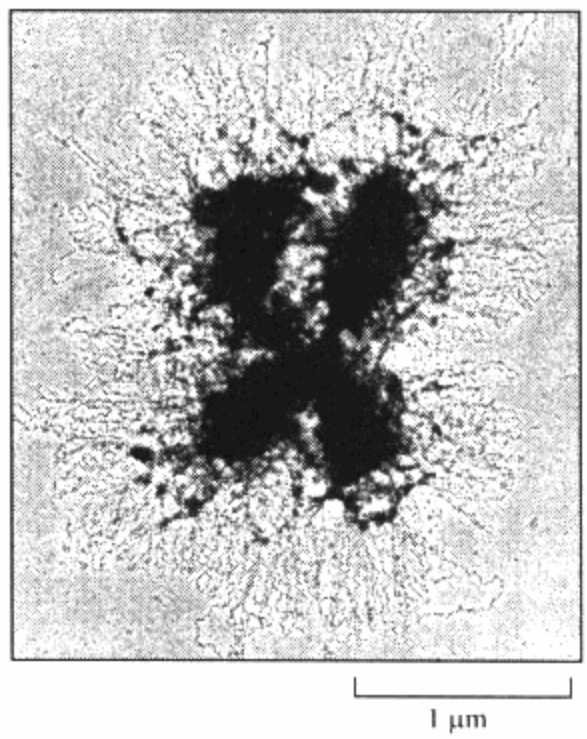


图 4-54　有丝分裂染色体电镜显微照片

该染色体（来自一种昆虫）经过处理，以显现出从染色单体中心骨架放射状伸出的染色体纤维环。显微照片支持了染色体折叠成一系列环状区域的观点（图 4-55）（经 Uli Laemmli 许可）

染色体在有丝分裂间期的压缩是一个高度组织和动态变化的过程，它至少有两个重要的目的。第一，当完全凝聚时（在中期），姐妹染色单体彼此解开，并行排列。因此，当有丝分裂装置开始将它们拉开时，姐妹染色单体就容易分离。第二，凝聚后的染色体在被拉牵引到两个子细胞中时，能保护较脆弱的 DNA 分子不发生断裂。

间期染色体凝聚成有丝分裂染色体的过程发生在 M 期，并且它与细胞周期的进程紧密相连（有关内容可见第 17 章和第 18 章的详细讨论）。该过程需要一类蛋白质，叫做凝聚蛋白，它利用 ATP 水解释放的能量，驱动每条间期染色体的缠绕，从而形成一条有丝分裂染色体。凝聚蛋白是含有 SMC 蛋白的巨大蛋白质复合体：长二聚体蛋白质分子固定在中间，两端是球状区域，由它结合 DNA 并水解 ATP（图 4-56）。如果结合上纯化的 DNA，凝聚蛋白就利用 ATP 水解释放的能量在 DNA 中形成巨大右旋的环。图4-56(C)中所示的盘绕模型基于这个事实：凝聚蛋白是有丝分裂染色体的主要结构成分，每 10 000 个核苷酸的有丝分裂 DNA 存在 1 分子的凝聚蛋白，尽管还不清楚它们如何作用于染色质。

每条有丝分裂染色体都含有巨大结构域这种特征模式

前面提到过，展示处于有丝分裂期的 46 条人类染色体的图称为人类核型。如果用如吉姆萨这样的染料染色后，有丝分裂染色体显示出显著且可再生的带型，它们排列在每条染色体上，如图 4-11 所示。这些带与前面描述的昆虫多线染色体带无关，后者相对于间期染色质较小的区

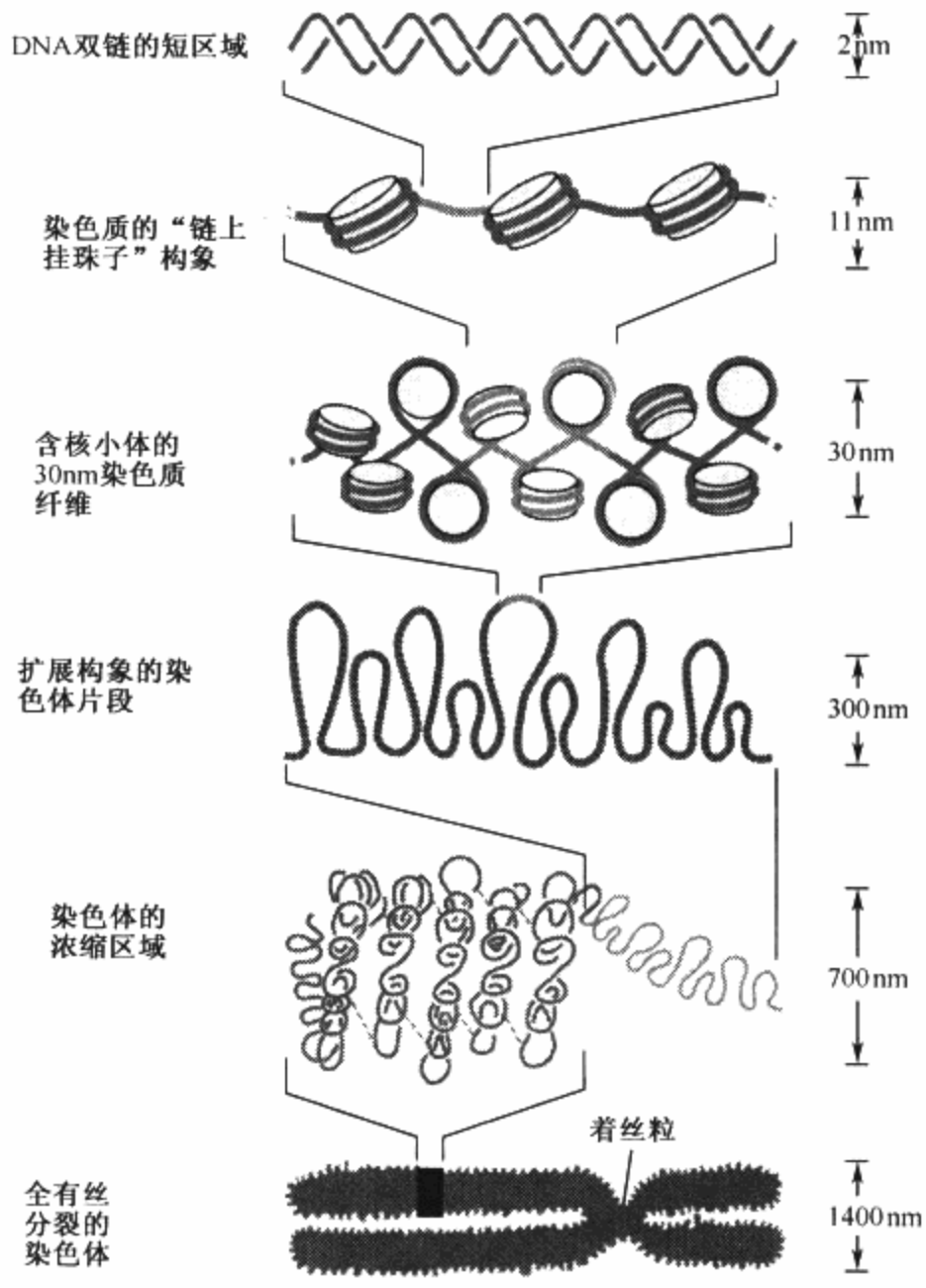


图 4-55　染色质包装

该模型显示了推测产生高度凝聚有丝分裂染色体的众多染色质包装水平中的一些

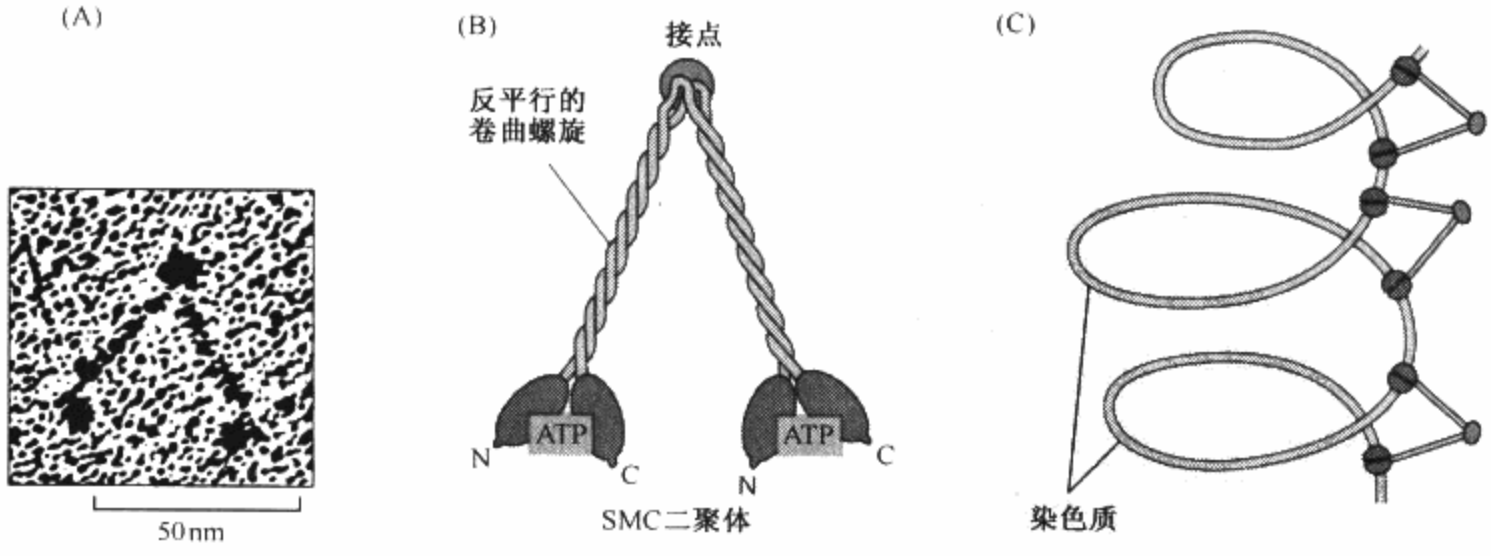


图 4-56　凝聚蛋白中的 SMC 蛋白

（A）纯化 SMC 二聚体的电镜照片。（B）SMC 二聚体的结构。该蛋白质中间长区域是反向平行卷曲螺旋（图 3-11），中间有一个柔性的绞索，如（A）中电镜照片所示。（C）说明凝聚蛋白中的 SMC 蛋白压缩染色质的可能模型。实际上，SMC 蛋白是这种大得多的凝聚蛋白复合体的组分。有人提出在细胞中，凝聚蛋白将长串的环状染色质区域盘绕起来（图 4-55）。通过这种方式，凝聚蛋白可以形成一个结构框架，在细胞周期 M 期中维持处于高度组织状态的 DNA

［(A)经 H. P. Erickson 许可；(B) 和 (C) 摘自 T. Hirano，*Genes Dev.* 13：11～19，1999］

域而言。在人类有丝分裂的染色体中，所有的染色质都被凝聚，这些带代表染料的选择性结合。

在有丝分裂早期，人类染色体处于比中期凝聚程度更小的状态，通过对它们进行考察，已经能估计出整个单倍体基因组含有大约 2000 条可分辨的带。有丝分裂期间随着凝聚程度逐渐加深，这些带就会逐渐合并，且越来越厚。我们在前面看到，细胞遗传学家通常利用这些染色体带型来发现患者的遗传改变，如染色体倒位、易位以及其他类型的染色体重排（图 4-12）。

图 4-57　最大的人类染色体(1 号染色体)与黑猩猩、大猩猩、猩猩的最大染色体的吉姆萨带型比较

比较所有这些染色体的染色模式后表明，人类染色体与黑猩猩的紧密相关，与大猩猩的其次，与猩猩的染色体相关性最远（摘自 M. W. Srickberger, Evolution, 3rd edn. Sudbury, MA: Johes & Bartlett Publishers, 2000）

从人类到果蝇，在各种各样的物种里都能找到有丝分裂染色体带。而且，染色体的带型经历长期的进化后仍然保持不变。例如，每条人类染色体在黑猩猩、大猩猩、猩猩的染色体中都有一条清晰可辨的染色体，它们具有几乎相同的带型——尽管仍存在明显的差别，例如，染色体融合使人类染色体数目为 46 条，而不是像猿那样有 48 条（图 4-57）。这表明染色体组织成大的结构域可能对染色体功能很重要。

在图 4-11 中，即使最窄的带也可能含有至少 100 万个核苷酸对，这几乎与细菌基因组的大小相当。这些带似乎反映出染色体大致划分为 GC 含量不同的区域。人类基因组的核苷酸序列揭示出巨大的非随机序列区段的存在（有些多于 10^7 核苷酸对），其 GC 含量显著高于或低于整个基因组 41%的平均含量。这些区段与中期染色体带型大致相关。例如，吉姆萨染色深的条带（所谓的 G 带）与低 GC 含量相关，而染色浅的条带（R 带）则对应于 GC 含量比平均值高的 DNA 序列。

一般来讲，基因组富含 GC 的区域基因密度较高，特别是“持家基因”，它们实际上是在所有细胞类型中都表达的基因。在这些观察的基础上，有人提出带型可能与基因表达有关。也许，染色体分化为 G 带和 R 带反映了某些细微差别，这由 GC 含量决定的，并以染色体环在这些区域包装的方式进行。如果这种观点是正确的，染色体的粗略划分可以看成是一种区室化作用形式，其中特定的参与细胞表达的细胞组分更集中在需要它们活性的 R 带。无论怎样，从这种讨论中我们可以明显地看到，我们还仅仅是刚开始瞥见大规模染色体组装的原理。

单个染色体占据间期细胞核内连续区域

本章前面我们看到真核生物染色体存在于细胞核内。然而细胞核仅仅是一个装着染色体的口袋；相反，这些染色体以及细胞核内的其他组分（我们在后面的章节将会遇到），都是高度组织的。染色体在有丝分裂间期很活跃且难以看见，这期间它们在核内的组织方式自 19 世纪以来，一直令生物学家们着迷 。尽管现在我们的了解远未彻底，但我们确实已经知道了这些染色体组装的一些有趣的特征。

某种程度上的染色体有序性，来源于染色体在有丝分裂末期时常具有的构型。就在细胞分裂前，凝聚的染色体由附着到着丝粒上的微管向纺锤体两极牵引；因此，当染色体移动时，着丝粒

在前，而远臂（在端粒中终止）滞后。在有些细胞核内，染色体倾向于在整个间期都保持这种所谓的 Rabl 定向，其着丝粒朝向核的一极，而端粒则朝向相反的另一极（图 4-58 和图 4-59）。

大多数间期细胞内的染色体并不是处于 Rabl 定向；相反，着丝粒似乎分散在细胞核内。大多数细胞比上面说明的特殊细胞有更长的细胞间期，这有可能为它们的染色体提供时间来呈现不同的构象（图 4-59）。然而，每条间期染色体确实倾向于占据核内一个连续且相对较小

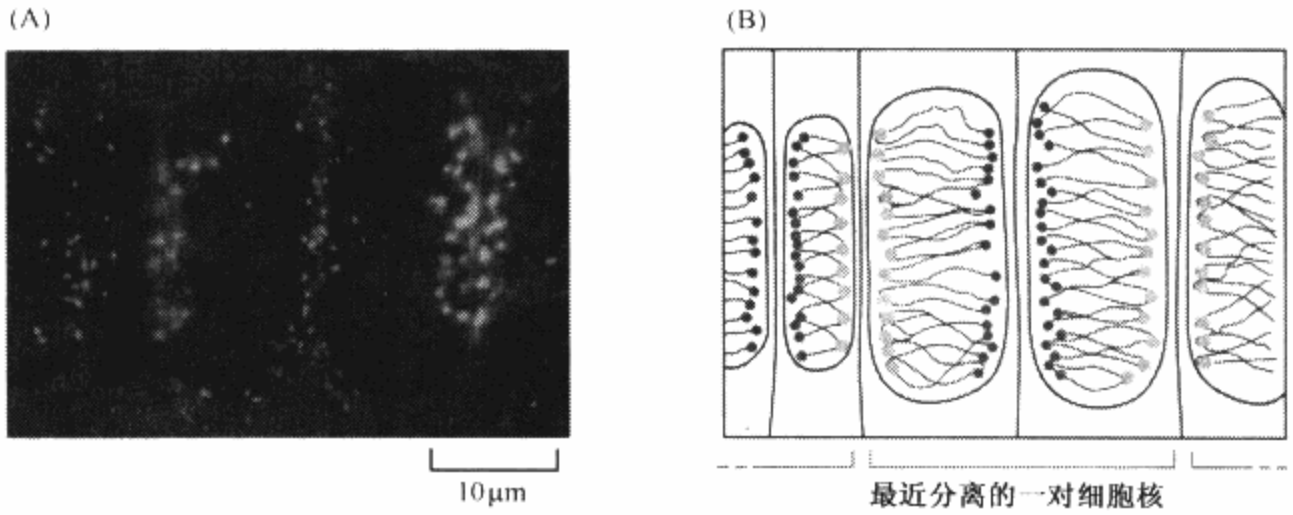


图 4-58　某些间期细胞核中发现的染色体极化方向

（A）迅速生长的植物根尖间期细胞核的荧光光学显微照片。利用已偶联上不同荧光染料的着丝粒以及端粒特异 DNA 序列进行原位杂交，将着丝粒染成绿色而端粒染成红色。（B）对（A）图的说明，图中显示位于 Rabl 定向的染色体，全部着丝粒朝向细胞核一边，而全部端粒指向相反的一边[(A)摘自 R. Abranches et al., *J. Cell Biol.* 143: 5～12, 1998. © The Rockefeller University Press]

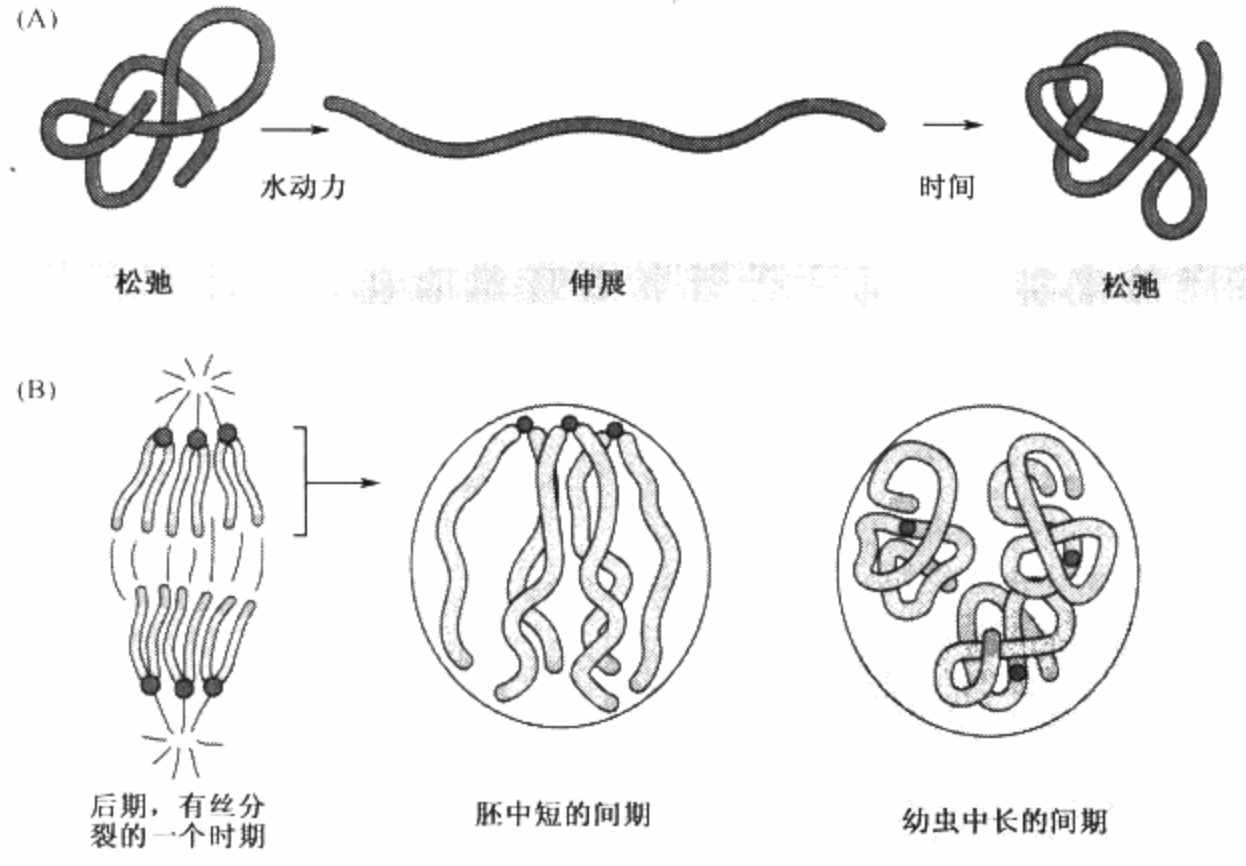


图 4-59　用多聚物来模拟间期染色体结构

（A）溶液中多聚物的行为。不存在外部施加的力量时，熵驱动一个长多聚物形成紧密的构象。如果多聚体受到剪切或者水动力的作用，它就变得伸展。但一旦这些作用力撤走，多聚体链又回到更加有利的紧密构象。（B）间期染色体的行为可能反映了相同的原理。如在果蝇胚胎中，有丝分裂每隔大约 10min 发生一次，在间隔短暂的间期里，染色体来不及从它们在有丝分裂期间运动诱导产生的 Rabl 方向上松弛下来。然而，在发育晚期，间期要长得多，染色体就有时间折叠起来。相同染色体不同区域之间的特异性联结可能强烈地影响了这种折叠（摘自 A. F. Dernburg et al., *Cell* 85: 745～759, 1996）

的区域；也就是说，不同的染色体彼此并不是大面积地缠绕着的（图 4-60）。

一种在核内组织这些染色体的方法可能是每条染色体的某一部分附着到核膜上（图 4-61）。例如，在许多细胞内，端粒似乎都以这种方式结合，但是在核内，染色体的确切位置并不固定。例如，同一组织中两个表面上相同的细胞靠得最近的染色体可以不相同。

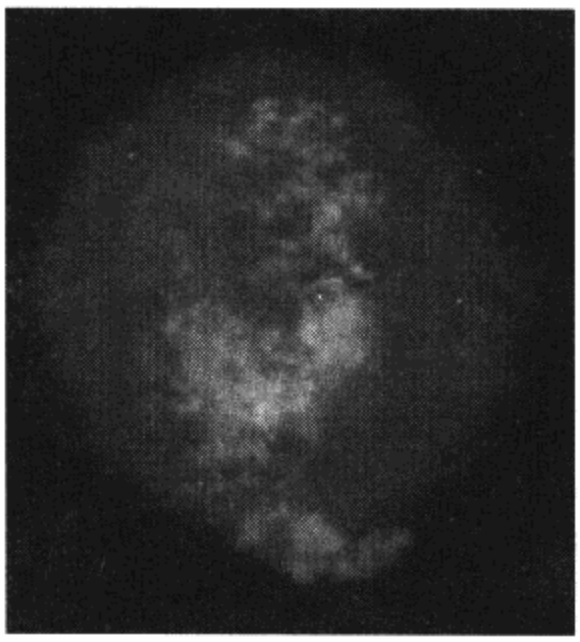


图 4-60　位于人类外周淋巴细胞中的两个间期染色体的选择性“染色”

这幅荧光光学显微照片表明，人类 18 号染色体（红色）的两份拷贝和 19 号染色体（青绿色）占据核内不连续区域（摘自 J. A. Croft et al., *J. Cell Biol.* 145：1119～1131，1999 © The Rockefeller University Press）

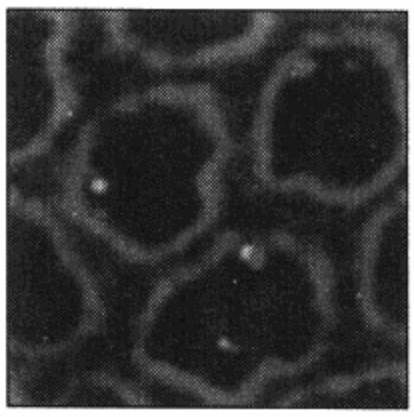


图 4-61　紧靠核被膜的间期染色体的特殊区域

这幅果蝇胚胎细胞核的高分辨率显微照片显示了 2 号染色体（黄色和品红色）紧靠核被膜的两个不同区域的位置，（用抗纤层的抗体染成绿色）。同一染色体的其他区域离核被膜更远一些（摘自 W. F. Marshall et al., *Mol. Boil. Cell* 7：825～842. 1996）

有些细胞生物学家认为，存在一种类似于细胞骨架的核内结构体系，染色体和其他细胞核组分在它上面组织起来。核基质或骨架，定义为经过一系列生物化学抽提过程以后核内剩余的不溶物质。组成它的蛋白质中有些可以证明与称作 SAR 或 MAR（骨架结合或基质结合区）的特殊 DNA 序列结合。据推测这些 DNA 序列构成染色体（图 4-44）环的基础，或者将染色体附着到核膜或核内其他结构上。通过这些染色体附着位点，基质可以帮助组织染色体、定位基因、调节基因表达和 DNA 复制。然而，是否生物学家分离到的基质就代表了完整细胞中存在的结构仍将不能肯定。

小结

染色体在有丝分裂间期通常解凝聚，这样就不能直接观察到它们。脊椎动物卵母细胞中特殊的灯刷染色体和昆虫巨大的分泌细胞中的多线染色体是显著的例外。这两种染色体的研究表明，每条染色体中冗长 DNA 分子分隔为大量以染色质环组织起来的不连续区域。每个环可能由折叠的 30nm 染色质纤维组成。当染色质环中包含的基因表达时，它就会解凝聚，使细胞装置能够容易地接触该 DNA。

常染色质占间期染色体的大部分，并且可能对应于 30nm 纤维的成环区域。然而常染色质被异染色质区间断，位于后者中的 30nm 纤维经过更高水平的包装，常常使它能抑制基因的表达。通常发现异染色质位于着丝粒两侧以及端粒附近，但它也存在于染色体的其他位置。间期染色体尽管远不如有丝分裂染色体那样凝聚，但在细胞核中占据不连续的区域；也就是

说，它们并非广泛地相互交织在一起。

在有丝分裂期间所有染色体呈现高度凝聚的构象，如果将它们进行特殊的染色，这些有丝分裂染色体就呈现出带状结构，使每条染色体都能清晰地加以辨别。这些带含有数百万个DNA核苷酸对，它们也反映了所知甚少的染色体结构的粗略异质性。

参考文献

一般文献

Hartwell L, Hood L, Goldberg ML et al. (2000) Genetics: from Genes to Genomes. Boston: McGraw Hill.

Lewin B (2000) Genes VII. Oxford: Oxford University Press.

Lodish H, Berk A, Zipursky SL et al. (2000) Molecular Cell Biology, 4th edn. New York: WH Freeman.

Wolfe A (1999) Chromatin: Structure and Function, 3rd edn. New York: Academic Press.

The Structure and Function of DNA

Avery OT, MacLeod CM & McCarty M (1944) Studies on the chemical nature of the substance inducing transformation of pneumococcal types. *J. Exp. Med.* 79, 137.

Meselson M & Stahl FW (1958) The replication of DNA in *E. coli*. *Proc. Natl Acad. Sci. USA* 44, 671–682.

Watson JD & Crick FHC (1953) Molecular structure of nucleic acids. A structure for deoxyribose nucleic acids. *Nature* 171, 737–738.

Chromosomal DNA and Its Packaging in the Chromatin Fiber

Aalfs JD & Kingston RE (2000) What does 'chromatin remodeling' mean? *Trends Biochem. Sci.* 25, 548–555.

Cairns BR (1998) Chromatin remodeling machines: similar motors, ulterior motives. *Trends Biochem. Sci.* 23, 20–25.

Carter NP (1994) Cytogenetic analysis by chromosome painting. *Cytometry* 18, 2–10.

Cheung P, Allis CD & Sassone-Corsi P (2000) Signaling to chromatin through histone modifications. *Cell* 103, 263–271.

Clark MS (1999) Comparative genomics: the key to understanding the Human Genome Project. *Bioessays* 21, 121–130.

DePamphilis ML (1999) Replication origins in metazoan chromosomes: fact or fiction? *Bioessays* 21, 5–16.

Dunham I, Shimizu N, Roe BA et al. (1999) The DNA sequence of human chromosome 22. *Nature* 402, 489–495.

Felsenfeld G (1985) DNA. *Sci. Am.* 253(4), 58–67.

Grunstein M (1992) Histones as regulators of genes. *Sci. Am.* 267(4), 68–74B.

International Human Genome Sequencing Consortium (2001) Initial sequencing and analysis of the human genome. *Nature* 409, 860–921.

Jenuwein T & Allis CD (2001) Translating the histone code. *Science* 293:1074–1080.

Kingston RE & Narlikar GJ (1999) ATP-dependent remodeling and acetylation as regulators of chromatin fluidity. *Genes Dev.* 13, 2339–2352.

Kornberg RD & Lorch Y (1999) Chromatin-modifying and -remodeling complexes. *Curr. Opin. Genet. Dev.* 9, 148–151.

Kornberg RD & Lorch Y (1999) Twenty-five years of the nucleosome, fundamental particle of the eukaryote chromosome. *Cell* 98, 285–294.

Luger K & Richmond TJ (1998) The histone tails of the nucleosome. *Curr. Opin. Genet. Dev.* 8, 140–146.

Luger K, Mader AW, Richmond RK et al. (1997) Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature* 389, 251–260.

McEachern MJ, Krauskopf A & Blackburn EH (2000) Telomeres and their control. *Annu. Rev. Genet.* 34, 331–358.

Ng HH & Bird A (2000) Histone deacetylases: silencers for hire. *Trends Biochem. Sci.* 25, 121–126.

O'Brien S, Menotti-Raymond M, Murphy W et al. (1999) The promise of comparative genomics in mammals. *Science* 286, 458–480.

Pidoux AL & Allshire RC (2000) Centromeres: getting a grip of chromosomes. *Curr. Opin. Cell Biol.* 12, 308–319.

Rhodes D (1997) Chromatin structure. The nucleosome core all wrapped up. *Nature* 389, 231–233.

Rice JC & Allis CD (2001) Histone methylation versus histone acetylation: new insights into epigenetic regulation. *Curr. Opin. Cell Biol.* 13, 263–273.

Ried T, Schrock E, Ning Y & Wienberg J (1998) Chromosome painting: a useful art. *Hum. Mol. Genet.* 7, 1619–1626.

Rubin GM (2001) Comparing species. *Nature* 409, 820–821.

Stewart A (1990) The functional organization of chromosomes and the nucleus—a special issue. *Trends Genet.* 6, 377–379.

Strahl BD & Allis CD (2000) The language of covalent histone modifications. *Nature* 403, 41–45.

Travers AA (1987) DNA bending and nucleosome positioning. *Trends Biochem. Sci.* 12, 108–112.

Wu J & Grunstein M (2000) 25 years after the nucleosome model: chromatin modifications. *Trends Biochem. Sci.* 25, 619–623.

The Global Structure of Chromosomes

Agard DA & Sedat JW (1983) Three-dimensional architecture of a polytene nucleus. *Nature* 302, 676–681.

Ashburner M, Chihara C, Meltzer P & Richards G (1974) Temporal control of puffing activity in polytene chromosomes. *Cold Spring Harbor Symp. Quant. Biol.* 38, 655–662.

Bickmore WA & Sumner AT (1989) Mammalian chromosome banding—an expression genome organization. *Trends Genet.* 5, 144–148.

Birchler JA, Bhadra MP & Bhadra U (2000) Making noise about silence: repression of repeated genes in animals. *Curr. Opin. Genet. Dev.* 10, 211–216.

Callan HG (1982) Lampbrush chromosomes. *Proc. Roy. Soc. Lond. Ser. B. (Biol.)* 214, 417–448.

Croft JA, Bridger JM, Boyle S et al. (1999) Differences in the localization and morphology of chromosomes in the human nucleus. *J. Cell Biol.* 145, 1119–1131.

Griffith JD, Comeau L, Rosenfield S et al. (1999) Mammalian telomeres end in a large duplex loop. *Cell* 97, 503–514.

Grunstein M (1997) Molecular model for telomeric heterochromatin in yeast. *Curr. Opin. Cell Biol.* 9, 383–387.

Hart CM & Laemmli UK (1998) Facilitation of chromatin dynamics by SARs. *Curr. Opin. Genet. Dev.* 8, 519–525.

Henikoff S (1990) Position-effect variegation after 60 years. *Trends Genet.* 6, 422–426.

Henikoff S (1998) Conspiracy of silence among repeated transgenes. *Bioessays* 20, 532–535.

Hirano T (1999) SMC-mediated chromosome mechanics: a conserved scheme from bacteria to vertebrates. *Genes Dev.* 13, 11–19.

Hirano T (2000) Chromosome cohesion, condensation, and separation. *Annu. Rev. Biochem.* 69, 115–144.

Lamond AI & Earnshaw WC (1998) Structure and function in the nucleus. *Science* 280, 547–553.

Lyko F & Paro R (1999) Chromosomal elements conferring epigenetic inheritance. *Bioessays* 21, 824–832.

Marsden M & Laemmli UK (1979) Metaphase chromosome structure: evidence for a radial loop model. *Cell* 17, 849–858.

Pluta AF, Mackay AM, Ainsztein AM et al. (1995) The centromere: hub of chromosomal activities. *Science* 270, 1591–1594.

Saitoh N, Goldberg I & Earnshaw WC (1995) The SMC proteins and the coming of age of the chromosome scaffold hypothesis. *Bioessays* 17, 759–766.

Spector DL (1993) Macromolecular domains within the cell nucleus. *Annu. Rev. Cell Biol.* 9, 265–315.

Thummel CS (1990) Puffs and gene regulation—molecular insights into the Drosophila ecdysone regulatory hierarchy. *Bioessays* 12, 561–568.

Weiler KS & Wakimoto BT (1995) Heterochromatin and gene expression in Drosophila. *Annu. Rev. Genet.* 29, 577–605.

Zhimulev IF (1998) Morphology and structure of polytene chromosomes. *Adv. Genet.* 37, 1–566.

5　DNA 复制、修复以及重组

DNA 序列的维持
DNA 复制机制
染色体上 DNA 复制的起始与完成
DNA 修复
一般性重组
位点特异性重组

细胞在一个混乱无序的宇宙中保持高度有序的能力，有赖于以 DNA 这种化学形式携带的大量遗传信息的精确复制。这个称为 DNA 复制的过程必须在细胞能够产生 2 个遗传上完全相同的子代细胞之前进行。保持有序性同样需要对这种遗传信息进行持续的监视和修复，因为细胞内的 DNA 不断受到多种因素的破坏，包括环境中的化学物质、辐射、热以及各种活性分子。这一章我们描述复制和修复细胞 DNA 的蛋白质装置。这些装置催化细胞内发生的一些最迅速、最精确的反应过程，它们的作用原理也清楚地证明了细胞化学的简练和有效性。

尽管单个细胞短期存活可以依赖于防止它的 DNA 发生改变，但是某个物种长期生存则要求序列在经历很多代后能发生变化。虽然细胞尽其所能保护它们的 DNA，DNA 仍然会偶尔发生改变。长此以往，这些变化就会产生物种进化过程中选择压力作用其上的遗传突变。

本章我们首先简要讨论 DNA 代代相传时发生的变化。然后我们讨论负责将这些变化维持在最小程度的机制——DNA 复制和 DNA 修复。最后，我们考虑细胞改变 DNA 序列的一些最有趣的方式，主要集中讨论 DNA 重组以及我们人类染色体上称为可移动元件的特殊 DNA 序列的移动。

DNA 序列的维持

尽管偶尔出现的遗传改变提高了物种长期存活的能力，但是个体的生存却需要遗传的稳定性。只在极为罕见的情况下，细胞的 DNA 维持程序会失灵，导致 DNA 发生永久性改变。这种改变称为突变（mutation），它如果发生在 DNA 序列中某个致命的位点，就会导致生物体的死亡。在讲述负责维持这种遗传稳定性的机制之前，我们先来简要讨论一下 DNA 序列带带相传时维持的精确性。

突变率极低

突变率（mutation rate）也就是 DNA 序列上发生可观察到的改变的比率，它可以直接通过在细菌，如大肠杆菌（我们肠道的一种寄生物）中进行实验来确定。在实验室条件下，大肠杆菌每 30min 分裂一次，不到一天的时间就可以由单个细胞获得一个很大的菌落——几十亿个。在这样一个菌落中，可以发现一小部分菌落在某个基因上发生了破坏性的突变，只要这个基因对于细菌的存活并不是必需的。例如，对于某个基因，细胞专门需要它以便利用半乳糖作为能量来源，如果将细胞放在存在一种不同的糖。例如，葡萄糖的条件下生长，那么就能确定这个基因的突变率（使用指示染料区分出突变的细胞）。损坏的基因部分要低于实际

突变率，因为许多突变都是沉默的（例如，有些情况改变了核苷酸密码，却没有改变它编码的氨基酸，或者即使改变了氨基酸也不会影响该基因编码蛋白质活性）。对这些基因沉默突变情况进行校正以后，估计编码平均大小蛋白质（约 10^3 个编码核苷酸对）在 10^6 代细菌细胞中只发生一次突变（该突变并不一定使蛋白质失活）。换句话讲，细菌每代细胞发生突变的概率大约是每 10^9 个核苷酸对中有 1 对会发生突变。

哺乳动物中生殖系的突变率更难直接测量，但可以进行间接的估计。一种方式就是比较多个物种中同一蛋白质的氨基酸序列。然后用任何 2 个物种间不同氨基酸部分进行比较。使用这种方法，可以计算氨基酸序列上发生的可遗传变化在生物体中固定下来之前，平均要经历的年数。这个值可以用来估计基因中产生单个稳定突变所需的平均年数，因为这种遗传变化的每一种通常反映该蛋白质编码基因的 DNA 序列单个核苷酸发生的改变。

这些计算几乎总是大大低估了实际突变率，因为许多突变会破坏蛋白质的功能，并因自然选择作用从物种中消失，也就是，通过含有这些突变生物的优先死亡。但是，也存在一个家族的蛋白质片段，它们的序列看起来不受影响，允许编码它们的基因累积突变，而不会被自然选择淘汰。这些是血纤肽，血纤肽是当凝血过程中纤维蛋白被激活形成纤维蛋白从纤维蛋白原释放出的 20 个氨基酸的片段。由于血纤肽的功能明显不依赖于它们的氨基酸序列，所以它们可以容忍任何氨基酸发生改变。血纤肽的序列比较表明，一个典型的含有 400 个氨基酸的蛋白质，大约每隔 200 000 年通过生殖系中单个氨基酸变化而发生一次随机改变。

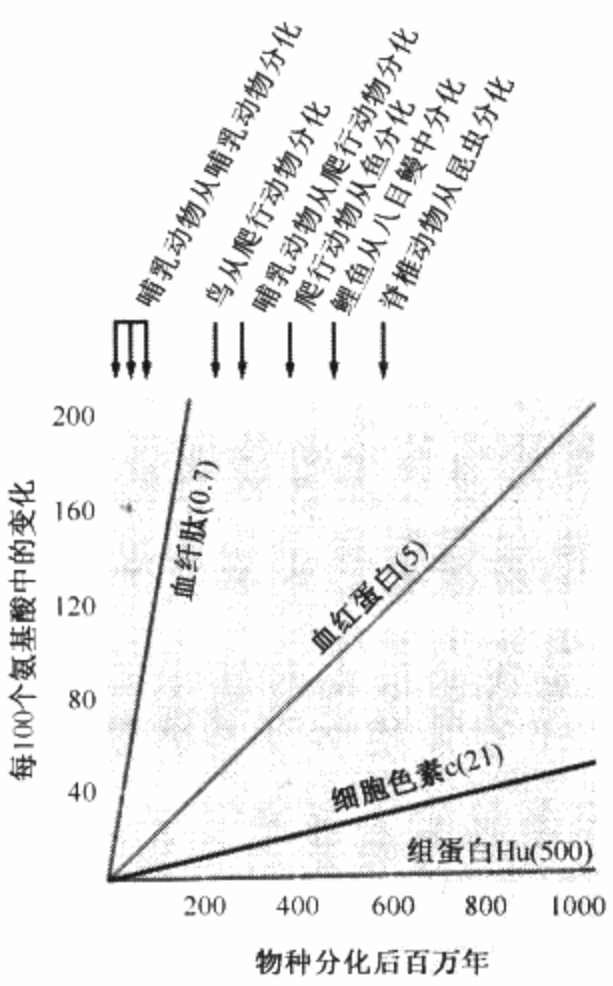


图 5-1　不同的蛋白质以不同的速度进化

血红蛋白、组蛋白 H4、细胞色素 c 和血纤肽中氨基酸改变速度的比较。前 3 种蛋白在进化中变化速度比血纤肽慢得多，圆括号内的数字表示蛋白质所含的每 100 个氨基酸中出现一个可接受的氨基酸改变平均要花费的百万年数。在决定每年的变化速率时，意识到 100 百万年前从共同的祖先趋异的 2 个物种经过 200 百万年的进化时间才彼此分开很重要

估计突变率的另一种方法就是利用 DNA 测序比较不同物种基因组中不含重要信息的区域内对应的核苷酸序列。这种比较估计出来的突变率与血纤肽的研究获得的突变率很接近。

大肠杆菌和人类之间在繁殖方式以及世代时间方面都相差很大。然而，如果将它们各自的突变率标准化为单轮的 DNA 复制，就发现它们各自突变率很相似了：每轮复制时，10^9 个核苷酸中大约有 1 个发生改变。

蛋白质中许多突变都是有害的，并被自然消除

如果将几对物种中特定蛋白质存在差异的氨基酸数目，相对于该对物种从共同祖先趋异进化的时间作图，结果是一条相当直的线：趋异进化时间越长，存在差异的氨基酸数目越大。为方便起见，这条直线的斜率用这种蛋白质的“单位进化时间”表述，也就是指 100 个氨基酸残基序列中出现 1 个氨基酸改变所需的平均时间。比较不同的蛋白质以后可以看到，它们每一个都显示出不同但很独特的进化概率（图 5-1）。

由于大部分 DNA 核苷酸经历大致相同概率的随机突变，不同蛋白质具有不同的突变率，必然反映一个氨基酸的改变对于每种蛋白质有害的可能性的不同。例如，从图 5-1 中数据我们可以估计到细胞色素

c 中每 7 个氨基酸改变中有 6 个是有害的，并且组蛋白 H4 中所有的氨基酸改变实际上都是有害的。携带这种有害突变的个体动物可能被自然选择群体淘汰。

据我们所知，低突变率对于生命是必需的

既然如此多的突变对于生物体有害，所以也就没有任何一个物种能够付得起代价，允许突变在它生殖细胞中以极高频率累积起来。尽管观察到的频率很低，人们仍然认为 10 倍高的突变率将任何生物体都能编码的必需蛋白质数目限制在 6000 左右。在这种情况下，比果蝇稍简单的生物体将停止进化。

多细胞生物的体细胞必须防止发生遗传改变，以保护每一生物个体，然而生殖细胞必须防止高频突变率，以维持物种的延续。体细胞内核苷酸的改变会产生各种各样的细胞，其中的一些通过自然选择，以牺牲生物体其余部分为代价，进行迅速地增殖。在极端的情况下，其结果就是不受控制的细胞增殖，称为癌，这种疾病导致了欧洲和北美洲中每年约 30%的死亡病例。这些死亡病例大部分是因为体细胞 DNA 序列中突变的累积（讨论见第 23 章）。突变频率的显著增加将会加快体细胞突变体的出现频率，因而可能灾难性地增加癌症的发生。由此可见，无论从拥有大量基因的物种的永久存活（生殖细胞稳定性），还是从防止体细胞突变导致癌症发生体细胞稳定性的角度讲，像人类这样的多细胞生物依赖于维持它的 DNA 序列的显著的高度忠实性。

在下一部分，我们将会看到，DNA得以成功地维持，不仅依赖于 DNA 序列复制并分配到子代细胞中的精确度，而且依赖于修复 DNA 基因辐射、化学物质或其他事件导致的突变的一系列酶的作用。

小结

在所有细胞中，DNA 序列都以高度的忠实性进行维持和复制。对于像细菌与人这样如此不同的生物而言，突变率也差不多是一样的，即 DNA 每次复制时，每 10^9 个核苷酸对中约有 1 个核苷酸发生改变。由于存在这种显著的精确性，人类基因组序列（大约 3×10^9 个核苷酸对）在每次细胞分裂时，仅仅约有 3 个核苷酸发生了变化。这使得大多数人能将遗传信息精确地代代相传，而且能够避免在体细胞中发生突变，防止产生癌症。

DNA 复制机制

所有的生物体在每一次细胞分裂之前，都必须极其精确地复制它们的 DNA。在这一部分，我们探讨一种复杂的“复制机器”，如何能够以每秒 1000 个核苷酸的速率复制 DNA 的同时，还能实现如此高的精确性。

碱基配对原则为 DNA 复制和修复提供了基础

第 1 章我们简单讨论过，DNA 模板化是指一股 DNA 链（或者 DNA 链中选定的部分）的核苷酸序列通过碱基配对（A 和 T 配对，G 和 C 配对），拷贝成一条互补的 DNA 序列的过程（图 5-2）。这个过程要求 DNA 模板链中每个核苷酸能被一个自由的（未聚合的）互补核苷酸识别，而且它还要求 DNA 双螺旋的两条链分开。这种分离使每个 DNA 碱基上的氢键供体和受体基团都暴露出来，以供与将要靠近合适的自由核苷酸配对，通过酶催化的聚合反应将它们排列起来形成一条新的 DNA 链。

第一个核苷酸聚合酶，**DNA 聚合酶**（DNA polymerase），于 1957 年被发现。作为该酶底物的自由核苷酸是双脱氧核糖核苷三磷酸，它们聚合形成 DNA 需要一条单链的 DNA 模板。图 5-3 和图 5-4 说明了这个反应逐步进行的机制。

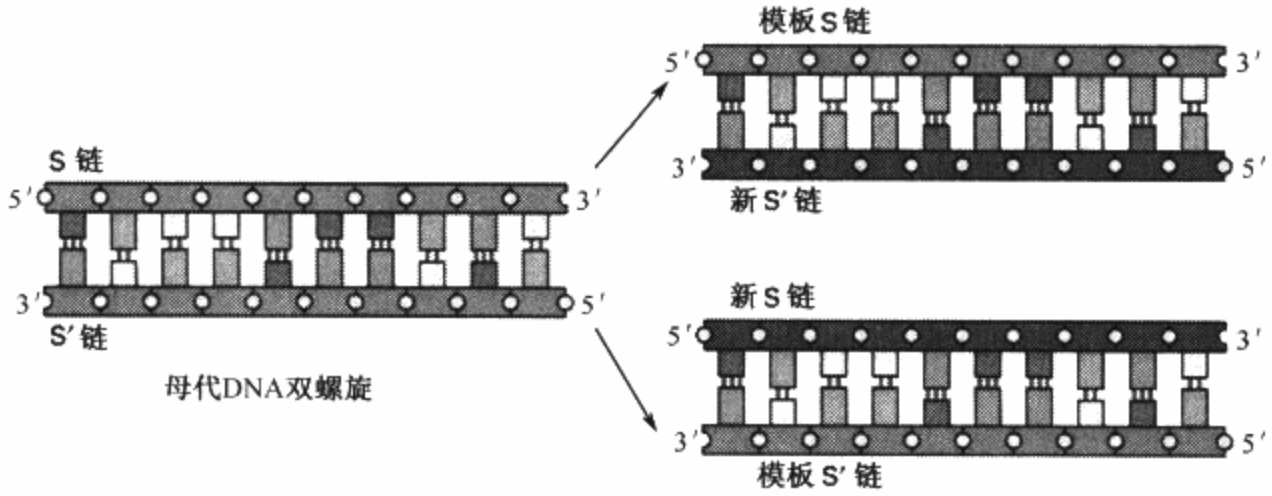


图 5-2　DNA 双螺旋作为自身复制的模板

因为核苷酸 A 只能成功地与 T 配对，G 只能与 C 配对，所以每股 DNA 都能用作模板，通过 DNA 碱基配对，指定它的互补链的核苷酸序列。通过这种方式，能够精确地拷贝出一条双链 DNA 分子

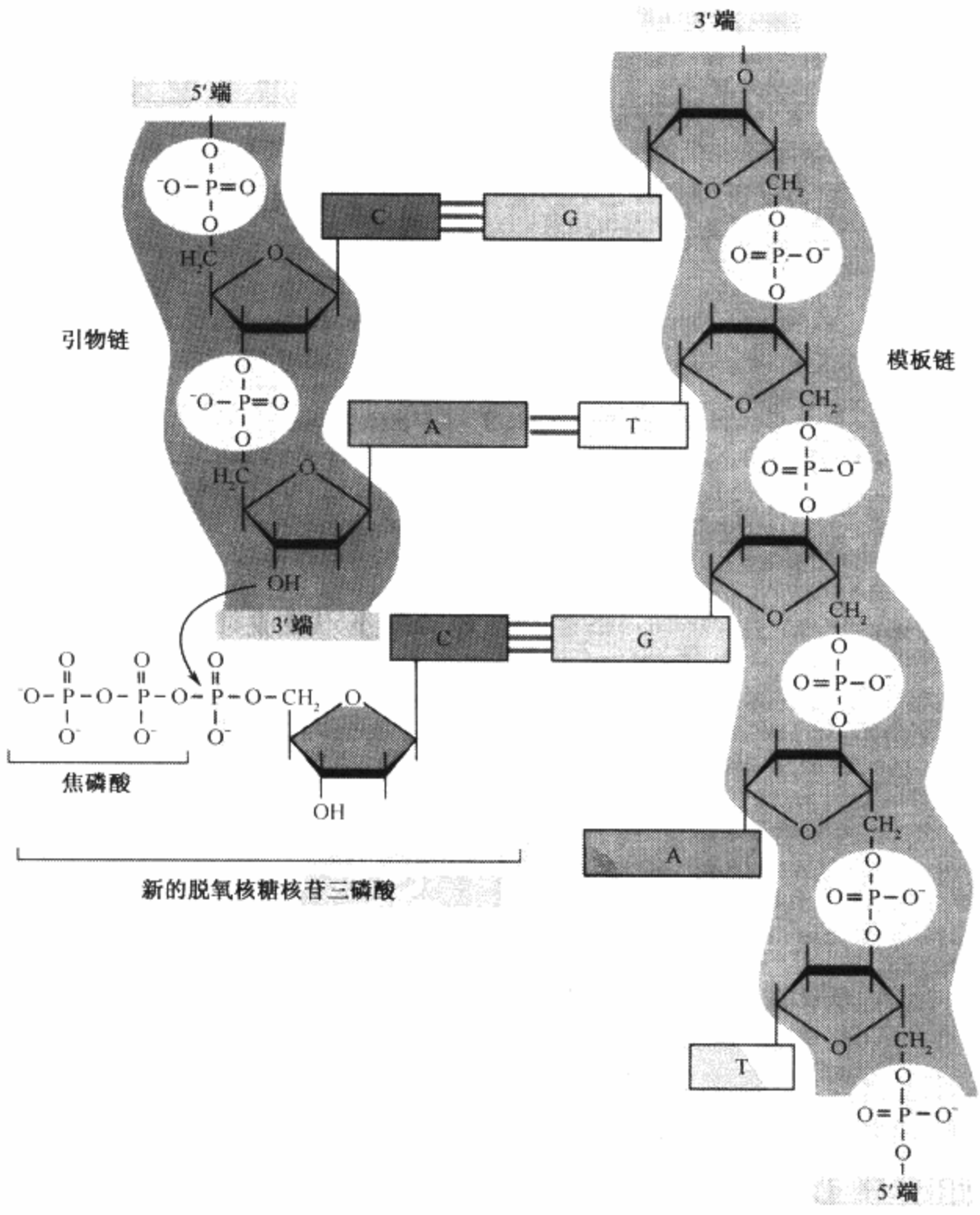


图 5-3　DNA 合成的化学

向一条多聚核苷酸链（引物链）的 3′端加入双脱氧核苷酸是 DNA 合成的基本反应。如图所示，即将进入的脱氧核苷酸三磷酸和已存在的 DNA 链（模板链）之间的碱基配对指导新的 DNA 链的形成，并使它具有互补的核苷酸序列

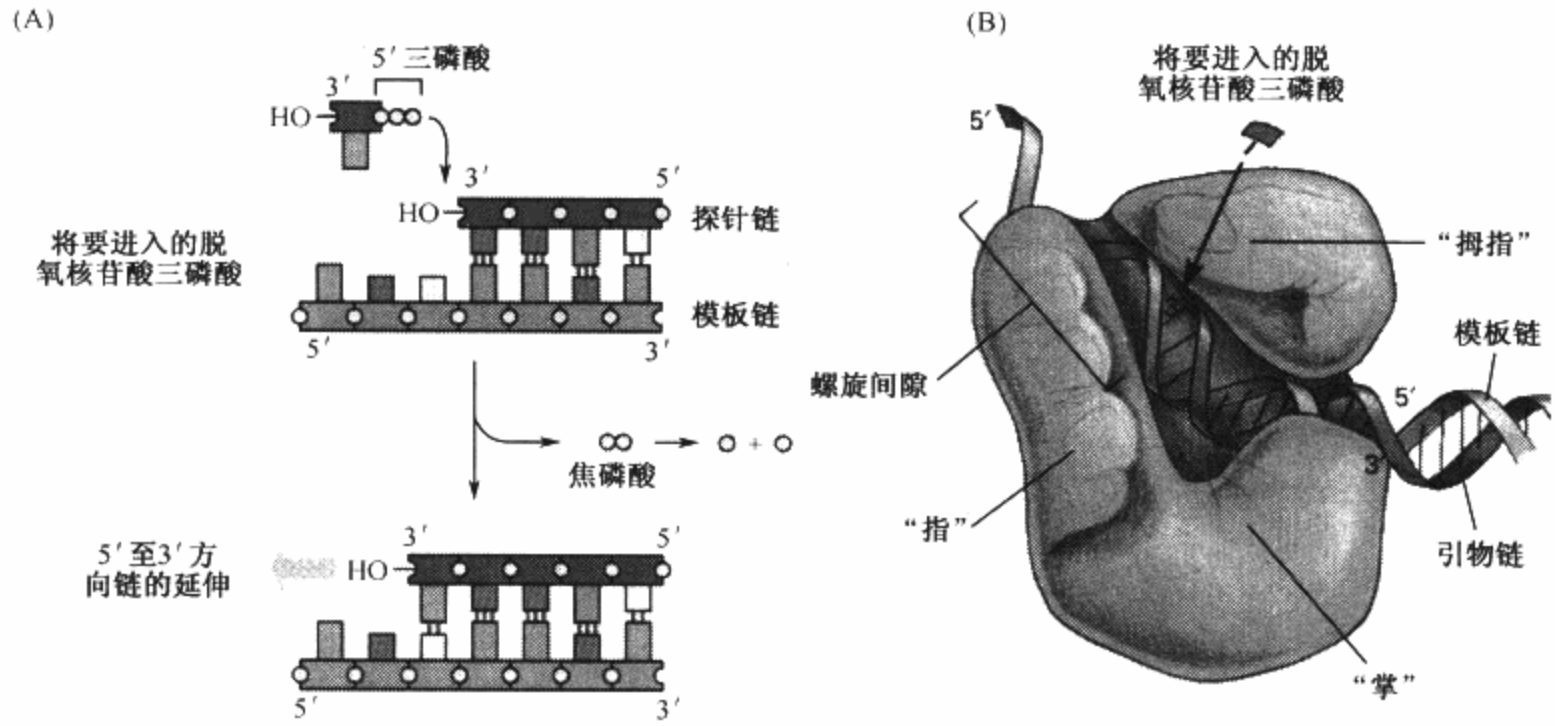


图 5-4　DNA 合成由 DNA 聚合酶催化

(A) DNA 聚合酶催化向一条多聚核苷酸链，引物链，的 3′-OH 端逐步加入脱氧核糖核苷酸，该引物链与另一条链互补配对。因此，正如前面的图中所示，新合成的 DNA 链以 5′→3′方向聚合。因为每个将要进入的脱氧核苷酸三磷酸必须与模板链配对，所以模板链决定 4 种脱氧核苷酸（A、C、G 或 T）中的哪个将被加入，该反应由焦磷酸释放及随后水解引起的大量的有利的自由能变化驱动。(B) 利用 X 射线晶体学确定的一个大肠杆菌 DNA 聚合酶分子的结构。大致来讲，它类似一个右手，其中手掌、手指和拇指紧抓着 DNA。这幅图说明的是在 DNA 修复中起作用的 DNA 聚合酶，但复制 DNA 的聚合酶具有类似的特征[(B) 摘自 L. S. Beese, V. Derbyshire, and T. A. Steitz, *Science* 260: 352～355, 1993]

DNA 复制叉是不对称的

细胞内 DNA 进行复制时，两股 DNA 链的每一条都作为模板，合成一条全新的链。由于分裂细胞的两个子细胞都继承了一个新的 DNA 双螺旋，它含有一条新链和一条旧链（图 5-5），所以说 DNA 聚合酶“半保留地”复制了 DNA 双螺旋。那么这是如何完成的呢？

19 世纪 60 年代早期，对所有正在复制的染色体进行分析后发现，存在一个复制的局部区域，它沿亲代 DNA 双螺旋不断向前移动。由于其 Y 形结构，这个活跃区域称为**复制叉**（replication fork）（图 5-6）。在复制叉位置，两条新的子代链都由一个含有 DNA 聚合酶的多酶复合体合成。

乍一看起来，DNA 复制的最简单机制似乎是：当复制叉从 DNA 分子一端移向另一端时，两条新链在复制叉位置不断生长，一个核苷酸接一个核苷酸。但是，由于 DNA 双螺旋中两股 DNA 链的反向平行（图 5-2），该机制要求一条子代链沿 5′→ 3′方向聚合，另一条沿 3′→ 5′方向聚合。这样的复制叉将需要两种不同的 DNA 聚合酶。一种 DNA 聚合酶催化由 5′→3′方向的聚合，这里进入的每一个脱氧核糖核

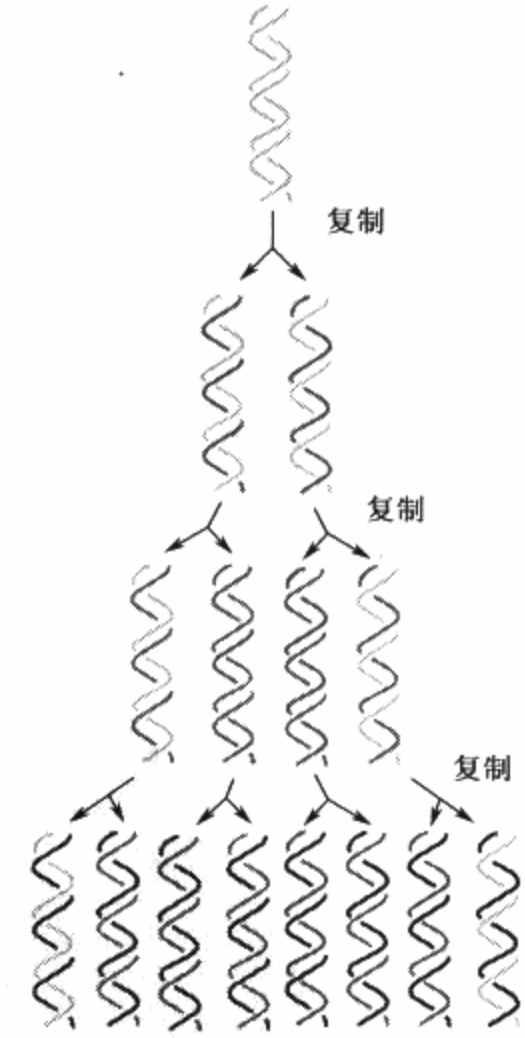


图 5-5　DNA 复制的半保留特性

每一轮复制中，两条 DNA 链每一条都用作模板合成互补的 DNA 链，因而原先的两条链经过许多的细胞世代仍然保留

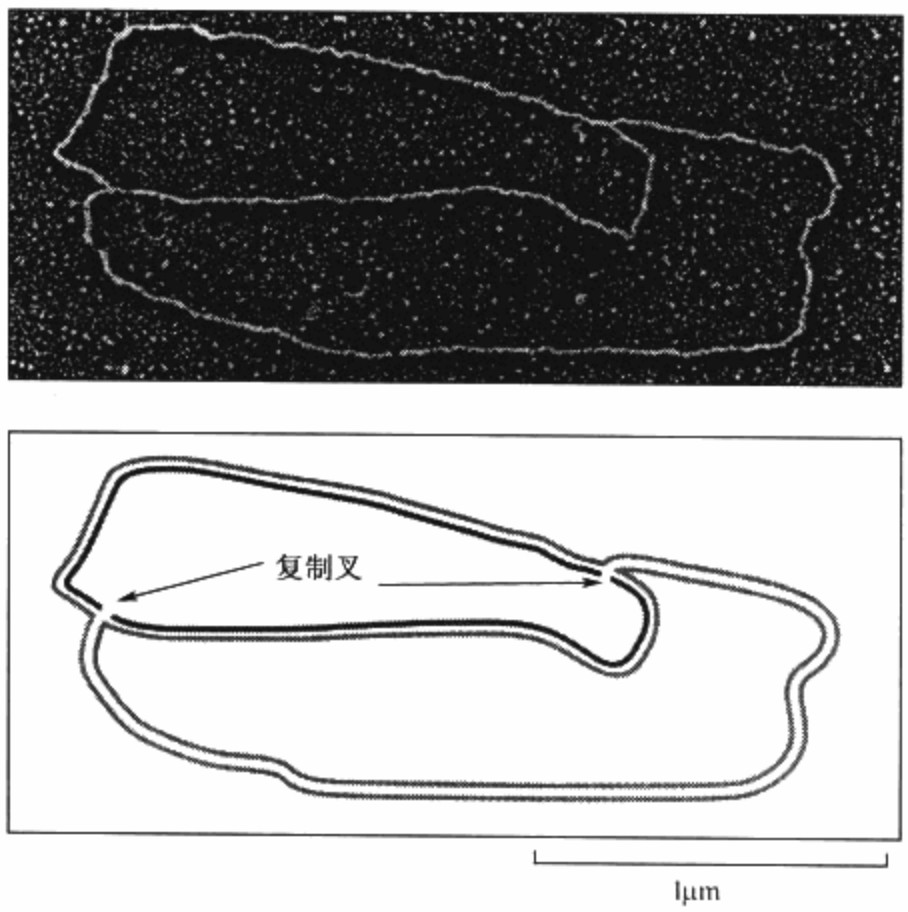


图 5-6　环状染色体上两个复制叉以相反的方向移动

DNA 复制活跃区域沿正在复制的 DNA 分子不断向前移动，形成一个称为复制叉的 Y 形 DNA 结构；每个 Y 的两臂是两个子 DNA 分子，Y 的茎秆是亲代 DNA 螺旋。在此图中，亲代链是橘黄色的；新合成的链是红色的（显微照片经 Terome Vinograd 允许后使用）

苷三磷酸携带一个激活的三磷酸以供自身的加成。另一种 DNA 聚合酶则催化由 3′→5′方向的聚合，通过所谓的“头部生长”进行工作，生长中的 DNA 链末端携带三磷酸活化能，以供接着进入的每一个三磷酸（图 5-7）。尽管头部生长的聚合反应在生物化学的其他地方存在（图 5-7），它却不在 DNA 合成中发生；还未发现 3′→5′的 DNA 聚合酶。

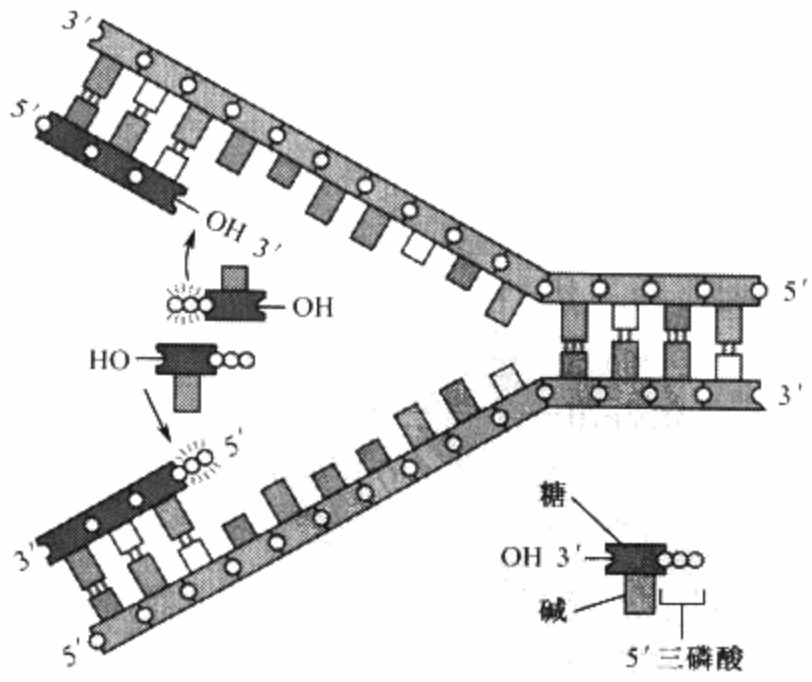


图 5-7　DNA 复制的一种不正确的模型

尽管它看起来是最简单的 DNA 复制可能的模型，但这里所说明的机制并不是细胞所利用的。在该图中，两条子 DNA 链将连续生长，利用两个末端磷酸的水解能（用红色射线突出黄色圆圈）将下一个核苷酸加入到每条链上。这将要求两条链都沿 5′→3′方向（下面）生长。还没有发现哪种酶催化 3′→5′方向的核苷酸聚合反应

那么 3′→5′的 DNA 链一般是如何生长的呢？19 世纪 60 年代得到的实验结果首先提示了这个问题的答案。研究者将高放射性的^{3}H-胸腺嘧啶加入到正在分裂的细胞中，过几秒钟，这样只有最新复制的 DNA（就是那些紧跟在复制叉后面的 DNA）被放射性标记上。这个实验揭示了在生长的复制叉位置，短暂存在一些 DNA 片段，它们长1000～2000 个核苷酸，现在通常称为冈崎片段。（类似的复制中间产物后来也在真核生物中发现，它们的长度仅为 100～200 个核苷酸。）实验表明，冈崎片段只从 5′→3′链方向聚合，当合成它们之后再连接起来形成长链 DNA。

因此，复制叉具有一个不对称的结构（图 5-8）。连续合成的子代 DNA 链称为**前导链**（leading strand）。前导链的合成比非连续合成的子代 DNA 链稍稍提前，这些非连续合成的子代 DNA 链称为**后随链**（lagging strand）。对于后随链，核苷酸聚合的方向与整体上 DNA 链生长的方向相反。后随链的合成被延迟是因为它必须等待前导链暴露冈崎片段合成的模板链。后随链以不连续的“倒缝”机制进行合成，意味着 DNA 复制只需要 5′→3′类型的 DNA 聚合酶。

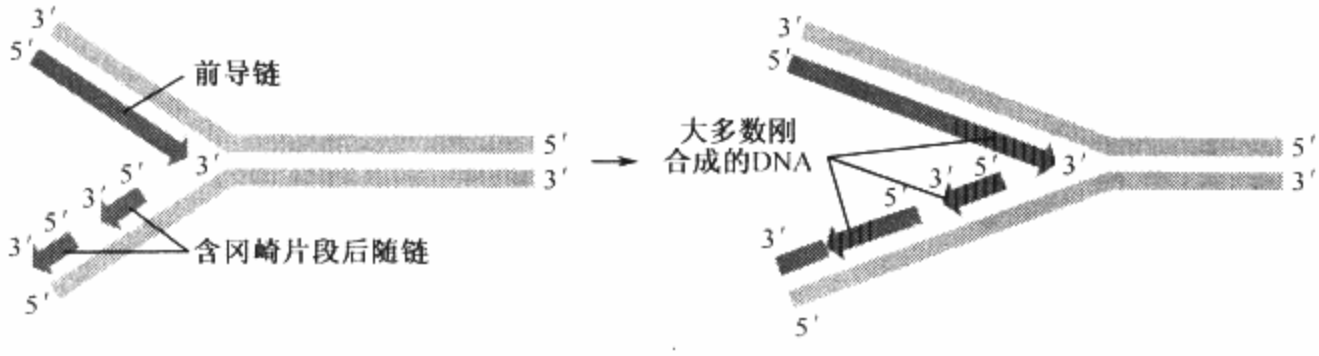


图 5-8　DNA 复制叉的结构

由于两条子代 DNA 链都以 5′→3′方向聚合，后随链上 DNA 合成起初必须先产生一系列短 DNA 分子，称作冈崎片段

DNA 复制的高度忠实性需要多种校对机制

本章开头我们讨论过，复制时拷贝 DNA 的忠实性可以达到每 10^{9} 个核苷酸只发生一次错误。这个忠实程度比我们基于互补碱基配对的精确性预测的要高得多。由此可见，标准的碱基配对（图 4-4）并不是唯一可能的校对机制。例如，一旦螺旋的几何形状发生小变化，就可以在 DNA 中的 G 和 C 之间形成氢键。此外，还存在 4 种 DNA 碱基的稀有互变异构体形式，其比率为 $1:10^{4}$ 或 $1:10^{5}$。这些互变异构体之间形成碱基错配却不改变螺旋的几何形状。例如，C 的稀有互变异构形式与 A 配对，而不与 G 配对。

当错配在即将进入的脱氧核糖核苷三磷酸与 DNA 模板之间发生时，如果 DNA 聚合酶没有采取特别的行动，那么这个错误核苷酸将总是整合到新合成的 DNA 链中，产生频繁的突变；然而，DNA 复制的高度忠实性不仅依赖于互补碱基配对，而且依赖于许多校对机制，它们连续发挥作用，以纠正早期可能出现的任何错误。

第一个校对步骤由 DNA 聚合酶执行，校对作用就在新的核苷酸进入生长链之前进行。我们关于这种机制的知识来自于几种不同聚合酶的研究，其中包括一种细菌病毒 T7 合成的 DNA 聚合酶，它在大肠杆菌内催化复制 DNA。对于移动中的聚合酶而言，正确配对的核苷酸能与模板正确地进行碱基配对。而且，在结合核苷酸后，但在核苷酸共价整合到生长链之前，聚合酶的构象必须发生变化。错配的核苷酸比正确的核苷酸在这一步更容易解离。因此，

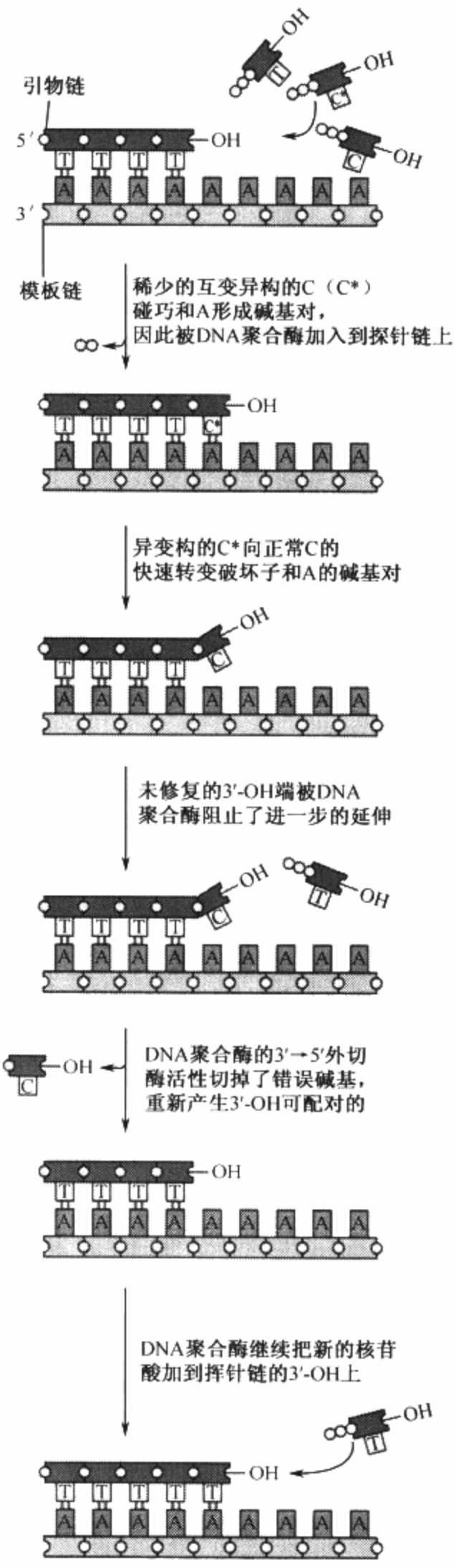


图 5-9　DNA 复制期间 DNA 聚合酶的核酸外切校对
本例中，胞嘧啶稀有且短暂存在的互变异构形式的掺入引起错配发生，用星号表示。但相同的校对机制适用于生长的 3′-OH 端任何的错误掺入

这一步使得 DNA 聚合酶在催化核苷酸的加成之前，能够对确切的碱基配对的几何结构进行“双重检验”。

在那些不常见的情况下，一旦不正确的核苷酸共价整合到生长链中，就会立即进行下一步的校错反应，称为外切核酸酶解校对。DNA 聚合酶并不能将两个核苷三磷酸连接起来，从而起始新的多聚核苷酸链的合成。相反，它们绝对需要一条碱基配对正确的引物链的 3′-OH 端，然后将其他的核苷酸结合上去（图 5-4）。那些在引物链的 3′-OH 端有错配（碱基不正确的配对）核苷酸的 DNA 分子并不是有效的模板，因为聚合酶不能延伸这样的链。DNA 聚合酶分子通过单独的一个催化位点来处理这种错配的引物链（或者是单独的亚基，或者是聚合酶分子单独的结构域，这取决于聚合酶）。这种 3′→5′的校正外切核酸酶将末端任何错配的残基切下来，直到足够的核苷酸被去除，以便重新生成一个能够引导 DNA 合成且碱基配对正确的 3′-OH 端。通过这种方式，DNA 聚合酶充当“自我校正”酶，当它沿 DNA 移动时，消除了自身的聚合错误（图 5-9 和图 5-10）。

对于 DNA 聚合酶自我校正的功能而言，一个碱基精确配对的引物末端的存在是极为关键的。很显然，在完全没有引物时这样一种 DNA 聚合酶能够起始合成，却又不失去辨别正确配对的 3′-OH 端是不可能的。相比之下，参与基因转录的 RNA 聚合酶不需要有效的核酸酶解的校对功能：合成 RNA 时出现的错误不会遗传给下一代，偶尔形成的缺陷 RNA 分子也没有长远的重要性。因此，RNA 聚合酶能够在没有引物存在的条件下起始合成多聚核酸链。

无论是在 RNA 合成，还是在 mRNA 序列翻译合成蛋白质序列的单独过程中，错误频率大概为 $1:10^4$。这个错误水平比 DNA 复制高出 100 000 倍，因为后者存在一系列校对过程，使得复制极为精确（表 5-1）。

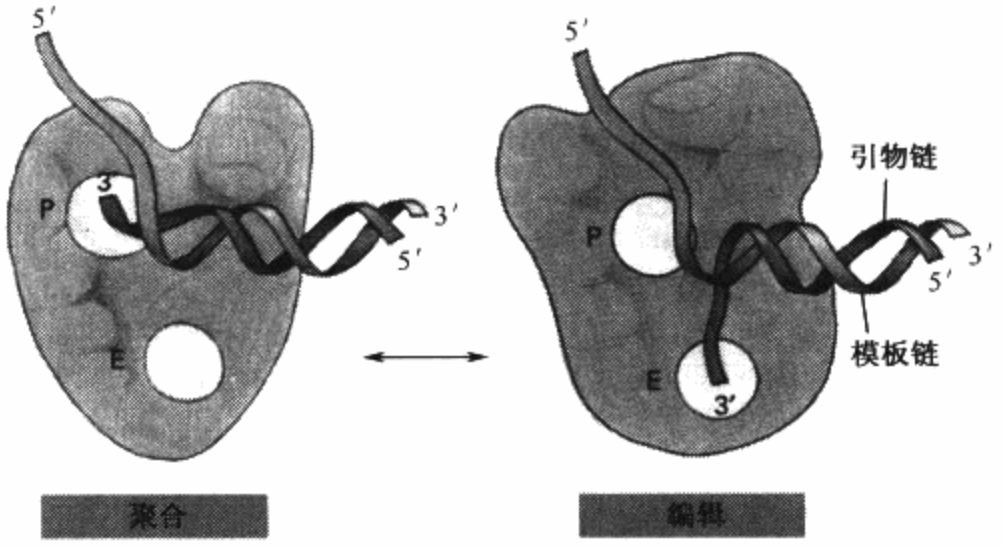


图 5-10　DNA 聚合酶的编辑

与 DNA 模板形成复合体的 DNA 聚合酶的结构轮廓，分别处于聚合模式（左）和编辑模式（右）。图中显示了核酸外切催化位点（E）和聚合催化位点（P）。研究者为了利用 X 射线晶体学确定这些结构，通过利用外切核酸结构域缺陷的突变将处于这两种状态的聚合酶“冷冻”，不能正常执行功能

表 5-1　产生高保真的 DNA 合成的 3 个步骤

复制步骤	每聚合一个核苷酸出现的错误
5′→3′聚合	1×10^{5}
3′→5′外切链指导的错配修复	1×10^{2}
总　　计	1×10^{9}

注：第三步（链指导的错配修复）将在本章后面的部分介绍。

只有以 5′→3′方向进行的 DNA 复制能够对错误有效地加以校正

对于精确性的要求也许能解释 DNA 复制为什么只以 5′→3′方向进行。如果存在一种 DNA 聚合酶，以 3′→5′方向加入脱氧核糖三磷酸，那么生长链的 5′端，而不是即将进入的单核苷酸，将携带活化的三磷酸。这种情况下，聚合时产生的错误就不能简单地加以水解掉，因为这样形成的光秃 5′端将立即中止 DNA 的合成（图 5-11）。由此可见，校正加入到 DNA 链 3′端的错配碱基比加入到 5′端的错配碱基容易得多。尽管 DNA 复制的机制（图 5-8），乍看起来比前面图 5-7 中描述的错误机制更加复杂，但是却更加精确，因为所有的 DNA 合成都以 5′→3′方向进行。

即便有这些保护措施防止 DNA 复制出现错误，但是 DNA 聚合酶本身也偶尔犯错误。然而，后面我们将会看到，细胞仍然有另外一种机会，通过一种称作链指导的错配修复的过程纠正这些错误。然而，在讨论这个机制之前，我们先描述一下在复制叉发挥作用的其他蛋白质。

一种特殊的核苷酸聚合酶在后随链上合成短 RNA 引物分子

对于前导链，仅仅在复制开始时需要一条特殊的引物：一旦形成复制叉以后，就持续提

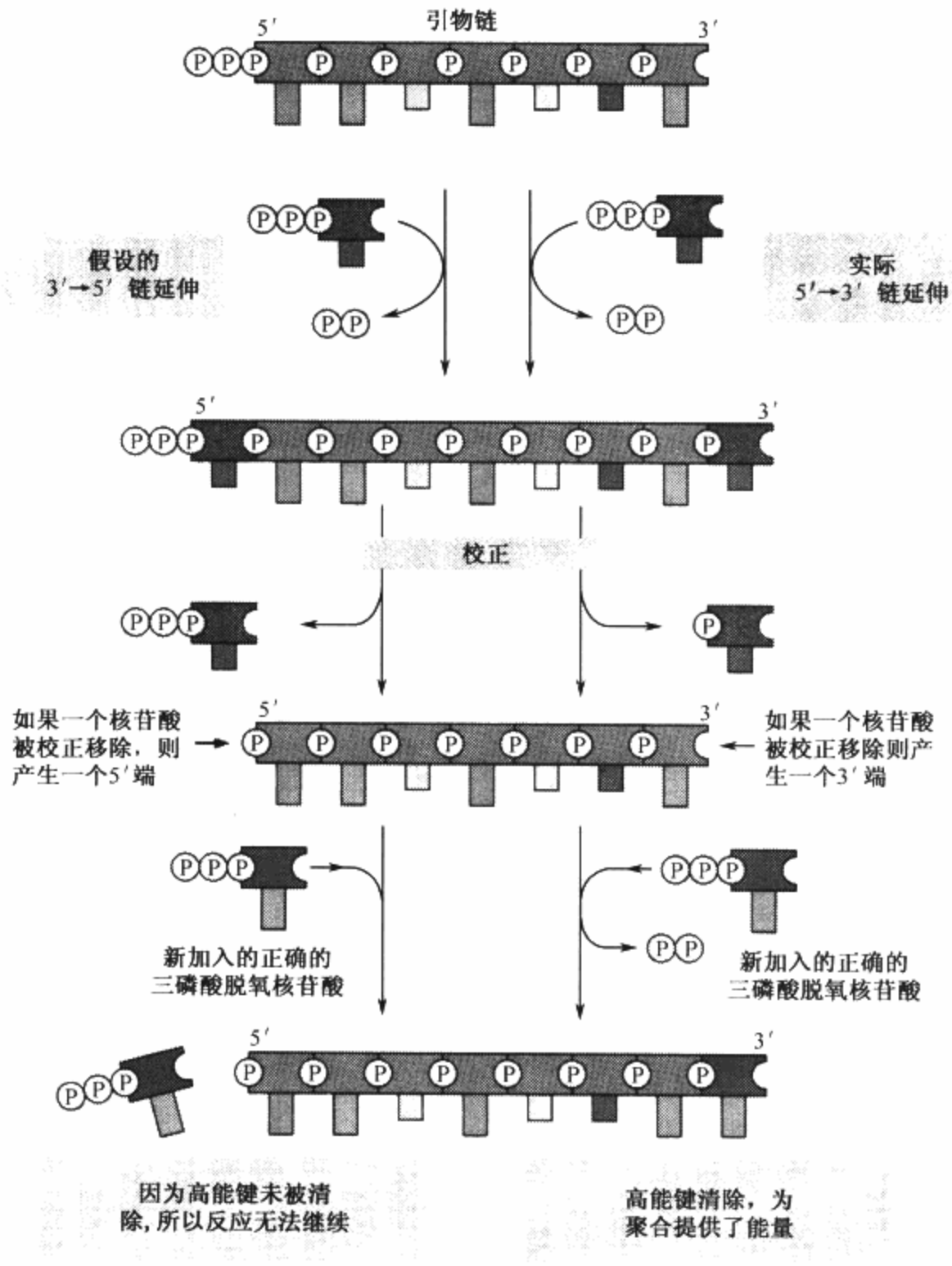


图 5-11　DNA 链沿 5′→3′方向生长的解释

右边显示，当聚合反应中的一个错误被核酸外切校对排除时，向 5′→3′方向的生长使链能够持续延伸（图 5-9）。相比之下，在假定的 3′→5′方向聚合体系中，左图所示，外切核酸校正进一步阻断链的延伸，为了方便起见，只显示了 DNA 双螺旋的引物链

供给 DNA 聚合酶一个碱基配对的引物末端，以便将新的核苷酸加上去。然而，在复制叉的后随链一侧，每一次 DNA 聚合酶完成冈崎片段合成以后（需要几秒钟），它就必须在模板链更远处起始合成一条全新的片段（图 5-8）。有一种特别的机制用来合成这种 DNA 聚合酶分子所需的碱基配对的引物链。这个机制涉及一种叫 **DNA 引发酶**（DNA primase）的酶的作用，DNA 引发酶利用核糖核苷三磷酸在后随链上合成短的 **RNA 引物**（RNA primer）（图 5-12）。在真核生物中，这些引物长度约为 10 个核苷酸，并且在后随链上每隔 100～200 个核苷酸合成一条。

RNA 的化学结构在第 1 章介绍过，在第 6 章做了详细的描述。这里，我们只需注意到

RNA 在结构上与 DNA 极为相似。一条 RNA 与 DNA 如果互补就可以形成碱基配对，生产一条 DNA/RNA 杂交双螺旋。因此，RNA 引物的合成与 DNA 合成都遵循相同的模板使用原则（图 1-5 和图 5-2）。

因为 RNA 引物在末端含有一个碱基正确配对的核苷酸，所有它能够在这一端由 DNA 聚合酶延长以起始冈崎片段的合成。当 DNA 聚合酶遇上与前面的冈崎片段的 5′端相连的 RNA 引物时，这条冈崎片段的合成就终止了。为了从后随链上合成的许多 DNA 片段生成一条连续的 DNA 链，有一种特殊的 DNA 修复系统迅速发挥功能，清除原来的 RNA 引物，并用 DNA 取代。然后一种 **DNA 连接酶**（DNA ligase）将新 DNA 片段的 3′端与前面合成的片段 5′端连接起来，完成整个复制过程（图 5-13 和图 5-14）。

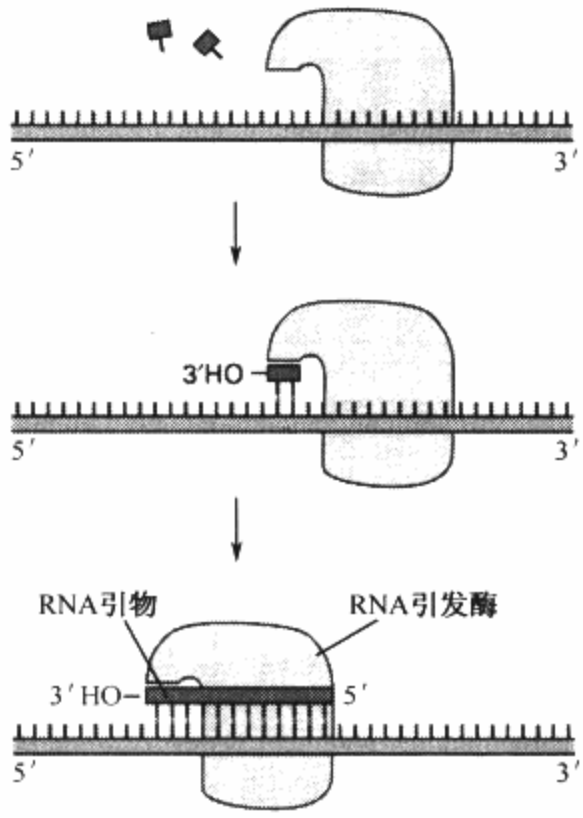


图 5-12　RNA 引物的合成

DNA 引发酶催化反应的简图。DNA 引发酶是一种利用 DNA 作为模板，在后随链上合成短 RNA 引物的酶。与 DNA 聚合酶不同，DNA 引发酶能将两个核糖核苷酸三磷酸连接起来，起始合成一条新的多聚核苷酸链。引发酶以 5′→3′方向合成一小段多聚核苷酸链，然后停下来，该引物的 3′端就能供 DNA 聚合酶利用

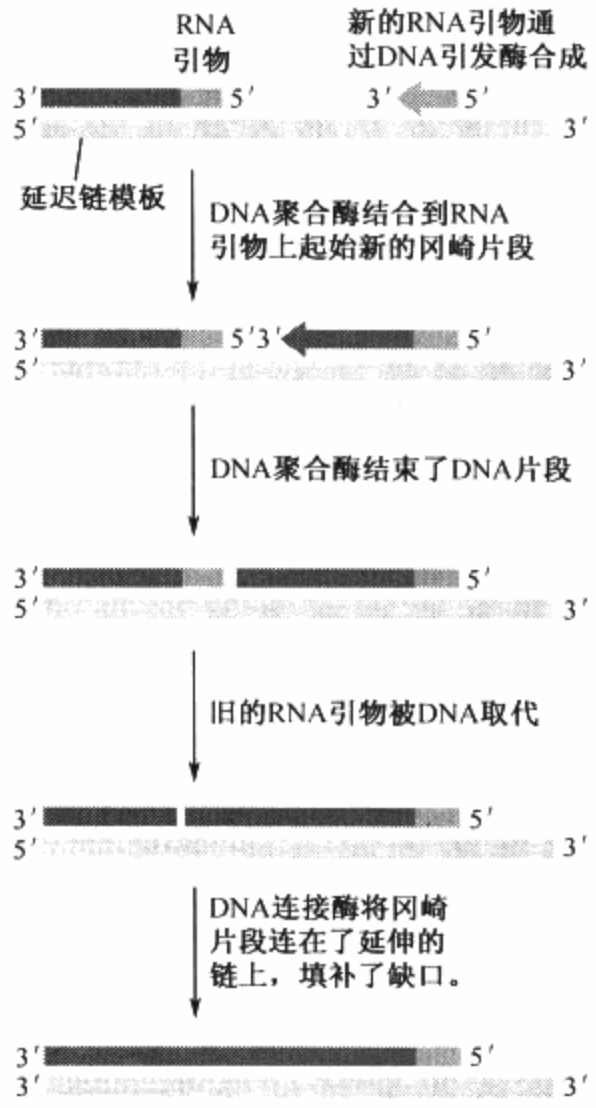


图 5-13　合成后随链上许多 DNA 片段中的一个

在真核生物中，RNA 引物在后随链上间隔大约 200 核苷酸合成，每个 RNA 引物长约 10 个核苷酸。这个引物由特殊的 DNA 修复酶（RNaseH）清除，它识别 DNA 螺旋中的 RNA 链，并将其片段化；作用产生的缺口由 DNA 聚合酶和 DNA 连接酶填补

为什么要优先使用可消除的 RNA 引物，而不使用不需要消除的 DNA 引物呢？认为自我校正的聚合酶不能重新起始链的合成的观点也暗示了它的逆否命题：能重新起始链的合成的聚合酶不能有效地进行自我校正。因此，任何引导合成冈崎片段的酶都必然会合成出一条相对不精确的拷贝（至少 10^5 个核苷酸中有一个错误）。即使这样的拷贝在终产物中保留下来，

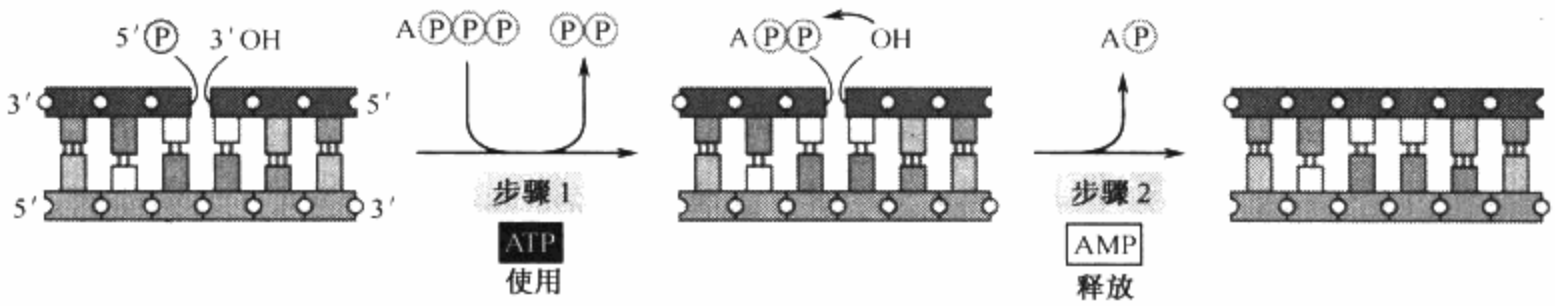


图 5-14　DNA 连接酶催化的反应

该酶封闭断裂的磷酸二酯键。如图所示，在形成新的键之前（第二步），DNA 连接酶利用一分子 ATP 活化切口的 5′端（第一步）。通过这种方式，能量上不利的切口封闭反应与能量上有利的 ATP 水解过程偶联起来

只占整个基因组不到 5%的比例（如每 200 个核苷酸的 DNA 片段中有 10 个核苷酸发生错配），其结果也会整体突变率大大增加。因而，看起来，进化将 RNA 而不是 DNA 用作引物似乎给细胞提供了一种强大的优势：引物中的核糖核苷酸自动将这些序列标记成“可疑拷贝”，以便对其进行有效的清除和取代。

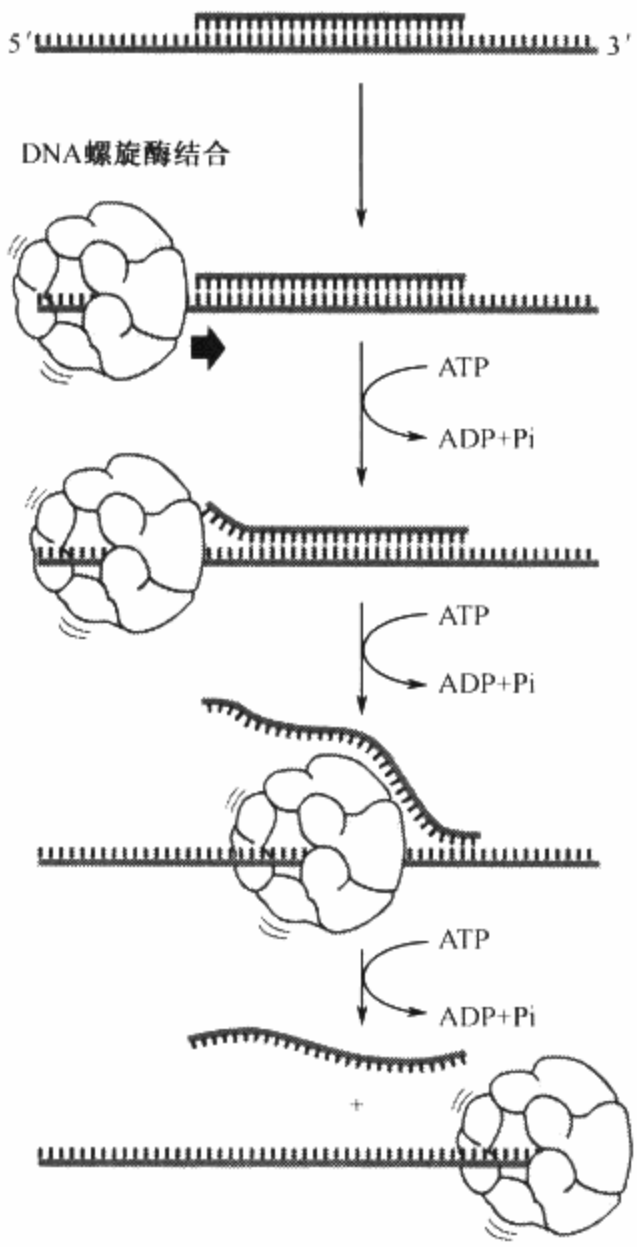


图 5-15　检验 DNA 解螺旋酶的实验分析

一小段 DNA 片段与一大段 DNA 单链退火形成一个 DNA 双螺旋区。随着解旋酶沿着 DNA 单链移动，双螺旋熔解，在需要解旋酶和 ATP 存在的反应中释放出小 DNA 片段。ATP 水解为解螺旋酶提供迅速地逐步移动的动力（图 3-76）

特殊蛋白质帮助打开复制叉前面的 DNA 双螺旋

为了使 DNA 合成方向推进，复制叉前面的 DNA 双螺旋必须打开，这样即将进入的双脱氧核糖核苷三磷酸才能与模板链形成碱基配对。然而，DNA 双螺旋在正常条件下很稳定；碱基对如此强有力地固定在其位置上，以至于在试管中需要接近沸水的温度才能将两条链分开。因此，只有当模板链与其互补链已经分离并暴露出来以后，DNA 聚合酶和 DNA 引发酶才能拷贝一条 DNA 双螺旋。此外，也需要其他的蛋白质帮助打开双螺旋，从而为 DNA 聚合酶的复制提供合适的单链模板。有两种蛋白质参与这个过程——DNA 解旋酶和单链 DNA 结合蛋白。

DNA 解旋酶（DNA helicase）最初是作为一种结合上单链 DNA 以后水解 ATP 的蛋白质分离出来。正如第 3 章所描述的，ATP 的水解能以可循环的方式改变蛋白质分子的形状，从而使得蛋白质做机械性重复的工作。DNA 解螺旋酶使用这一原理，让自身沿 DNA 单链迅速推进。当它们遇到双螺旋区域时，继续沿它们的链移动，从而以高达每秒 1000 核苷酸对的速率分开螺旋（图 5-15 和图 5-16）。

事实上，模板 DNA 双螺旋在复制叉位置的解旋可以由两种 DNA 解螺旋酶协同催化——一种沿

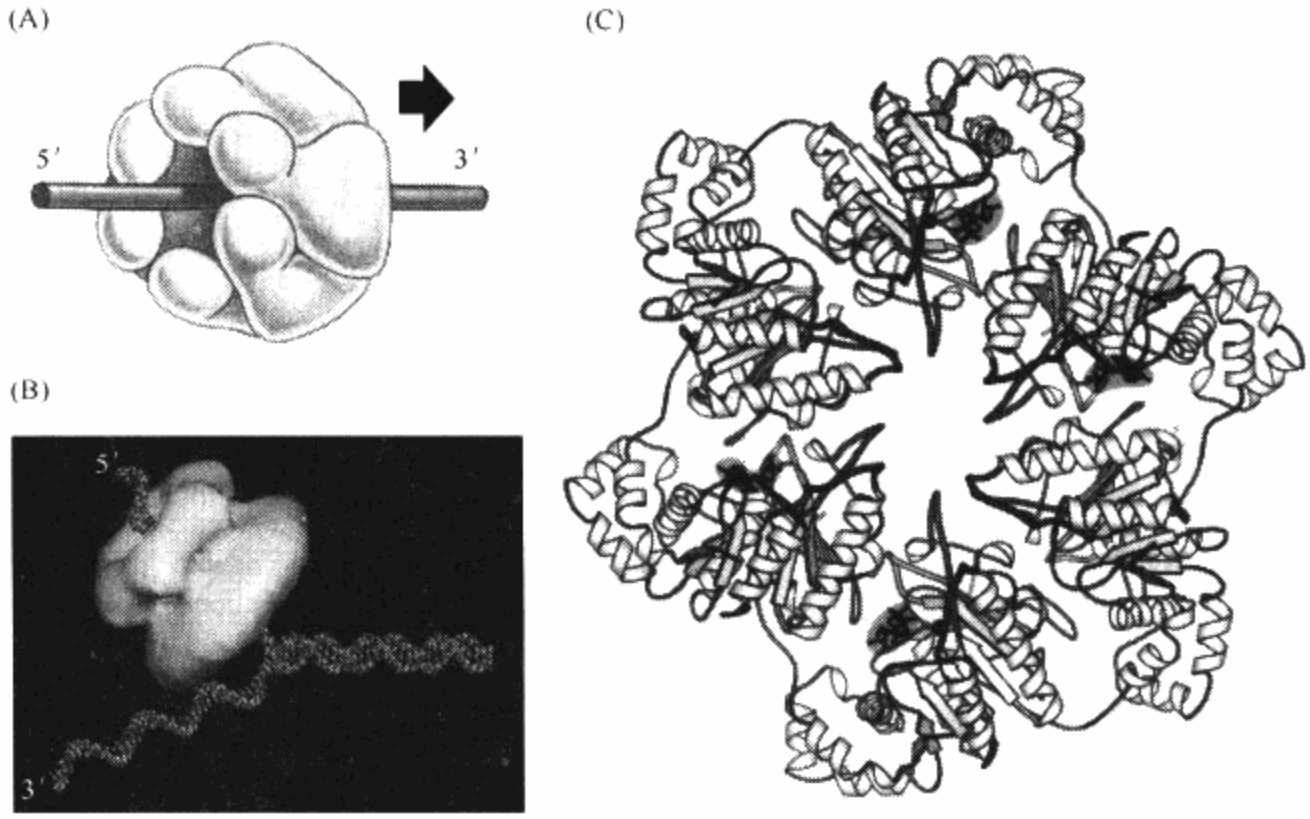


图 5-16　DNA 解螺旋酶的结构

（A）作为六聚体环蛋白的简图。（B）按比例显示 DNA 复制叉和解旋酶的简图。（C）X 射线衍射得到的噬菌体 T7 复制解螺旋酶的详细结构。6 个相同的亚基以一种有序的方式结合并水解 ATP，来推进该分子沿从中间孔穿过的 DNA 单链前进，红色表示结构内部结合的 ATP 分子[(B)经 Edward H. Egelman 允许使用；（C）摘自 M. R. Singleton et al.,*Cell* 101：589～600，2000. © elserier]

前导链模板运动，一种沿后随链运动。由于两条链的极性相反，这些解螺旋酶需要沿单链 DNA 向相反的方向移动，因此这些酶也就不同了。这两种类型的解螺旋酶都存在。在了解得最清楚的复制系统中，位于后随链模板上的解螺旋酶似乎发挥着主要作用，其原因不久就会弄清楚。

单链 DNA 结合（SSB）蛋白（single-strand DNA-binding protein），也称为螺旋去稳定蛋白，紧密而且协同地结合到暴露的单链 DNA 上，但并不覆盖碱基，因而使单链 DNA 可以作为模板使用。这些单链 DNA 结合蛋白不能直接打开一段 DNA 螺旋，但是它们能够帮助解旋酶稳定已开旋的单链构象。此外，它们协同地结合覆盖在单链 DNA 上以后，使该区域变得平直，从而能够防止短发卡状螺旋的形成，短发卡状螺旋在单链 DNA 内部容易形成（图5-17 和图 5-18）。这些短发卡状螺旋可以妨碍由 DNA 聚合酶催化的 DNA 合成。

移动的 DNA 聚合酶分子通过一个滑动的环保持与 DNA 的连接

依据不同的情况，大部分的 DNA 聚合酶分子从 DNA 模板上掉下来之前都只合成一小段核酸。从 DNA 分子上迅速解离的能力，使得后随链上刚刚完成一段冈崎片段合成的 DNA 聚合酶分子能够迅速地循环利用，以便在同一条链上开始下一段冈崎片段的合成。然而，要不是存在一种辅助蛋白质充当受调节的夹子的话，解离将使得聚合酶难以在复制叉位置合成长段的 DNA 链。当聚合酶移动时，该夹子牢固地将聚合酶限定在 DNA 上，而当聚合酶遇到前面 DNA 的双螺旋区域时，就立即释放它。

这个夹子是如何能够防止聚合酶解离的同时，又不会阻碍它沿 DNA 分子迅速移动的呢？应用 X 射线衍射得到的这个夹子蛋白质的三维结构表明它沿 DNA 螺旋形成了一个大环。这

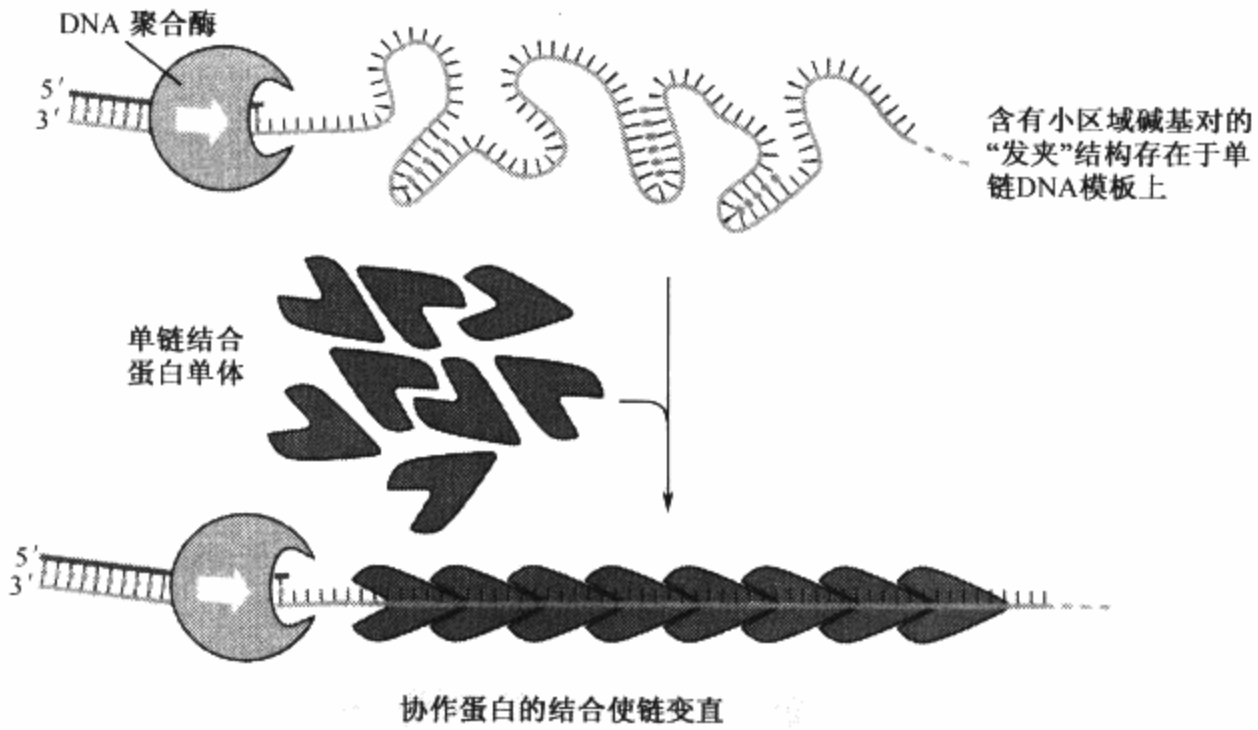


图 5-17　单链 DNA 结合蛋白（SSB 蛋白）对单链 DNA 结构的影响

因为每个蛋白质分子优先紧挨着结合先结合的分子，所以该蛋白质就在 DNA 单链上形成一长排。这种协同性结合使 DNA 模板变直，并促进了 DNA 聚合反应过程。小区域的互补核苷酸序列偶然配对形成上面所示裸露的单链 DNA 中的“发夹螺旋”；它们与通常在 RNA 分子中形成的短螺旋类似（图 1-6）

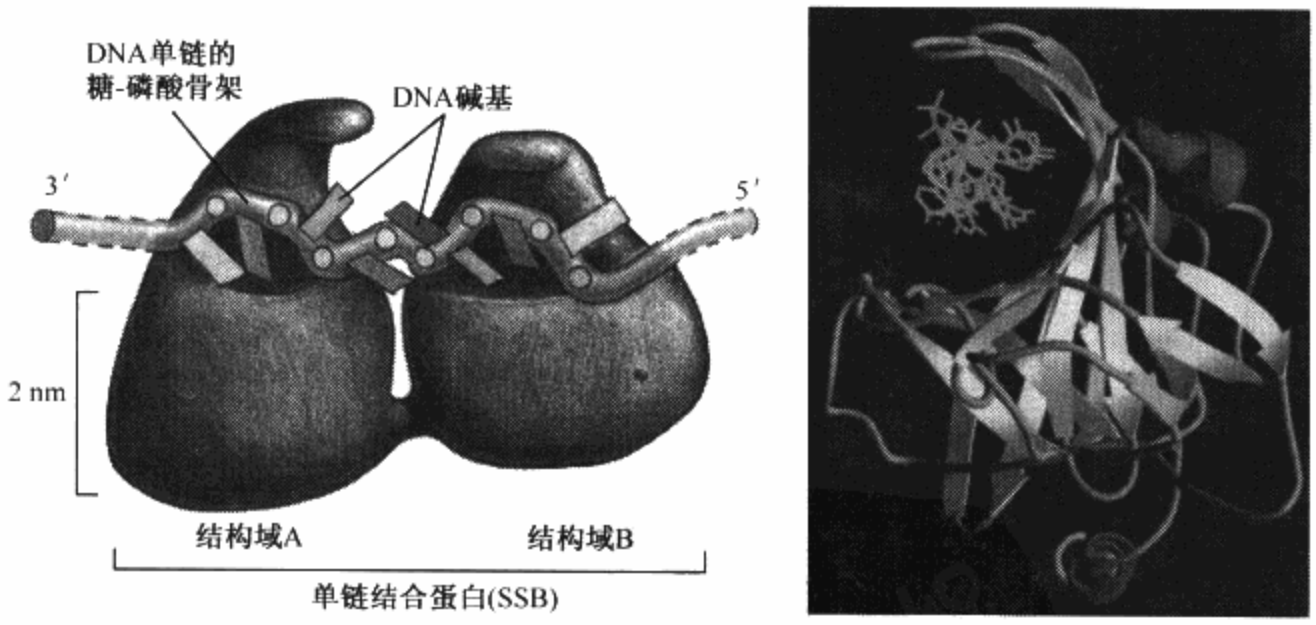


图 5-18　人类的与 DNA 结合的单链结合蛋白的结构

（A）RPA 蛋白的两个 DNA 结合结构域的前视图，该蛋白质覆盖总共 8 个核苷酸。注意该蛋白质-DNA 复合物中 DNA 碱基保持裸露。（B）三维结构图，其中 DNA 链（红色）从末端观察[(B)摘自 A. Bochkarev et al.,*Natune* 385：176～181，1997. © Macmillan Magazines ltd]

个环的一边结合到 DNA 聚合酶的后部，且整个环在聚合酶的移动时可沿 DNA 自由滑动。绕 DNA 组装这个夹子需要 ATP 的水解，这由特殊的蛋白质复合体，夹子装载器催化，它将夹子放置到引物-模板衔接点上时就会催化水解 ATP（图 5-19）。

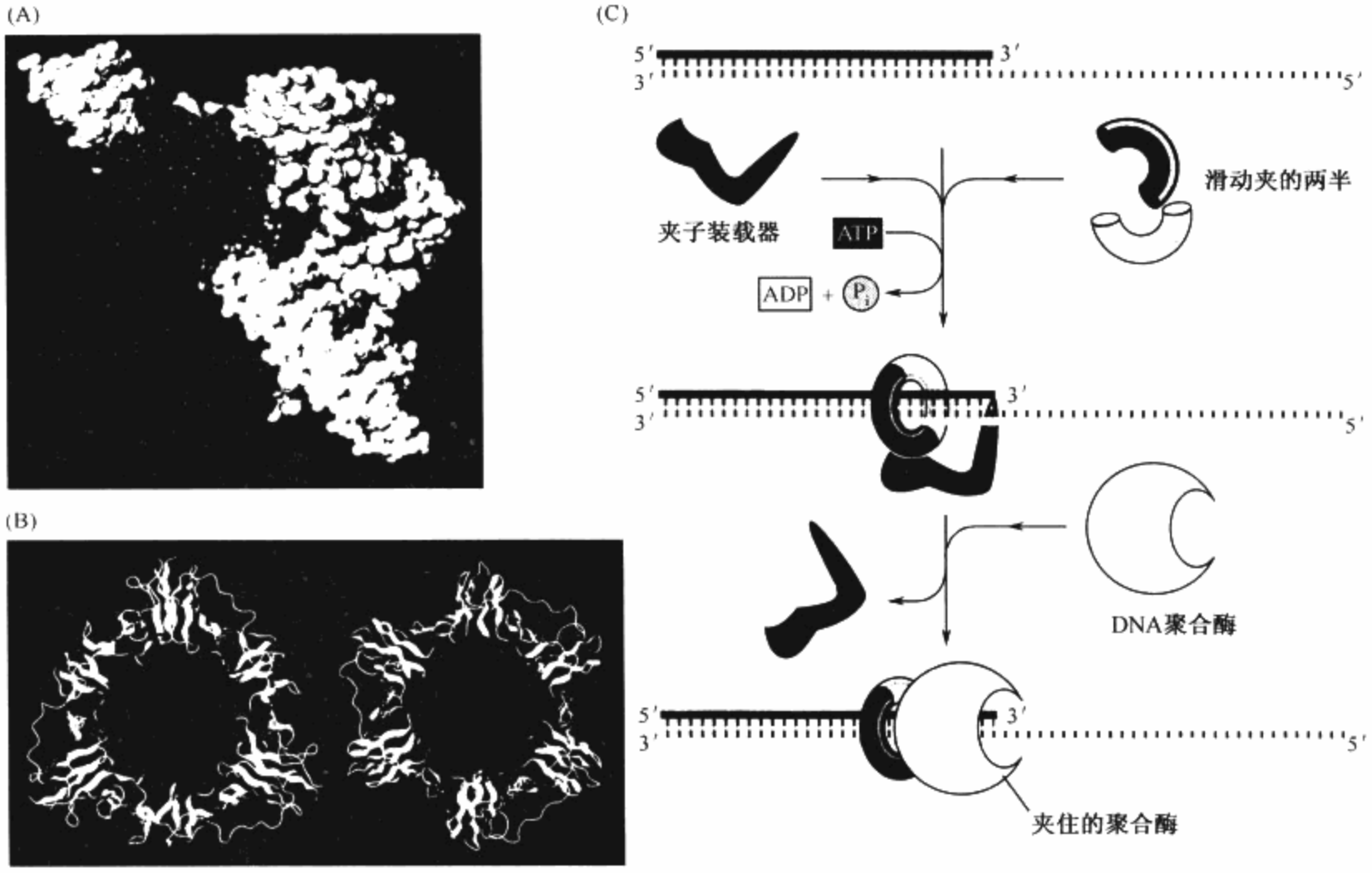


图 5-19　将 DNA 聚合酶保持在 DNA 上的可调节滑动夹

（A）利用 X 射线晶体学得到的来自大肠杆菌夹子蛋白质的结构，其中加入一个 DNA 螺旋表示该蛋白质如何正好环绕 DNA。（B）真核生物中存在的类似蛋白质，利用大肠杆菌滑动夹（左边）与人的 PCNA 蛋白（右边）进行比较加以说明。（C）显示夹子如何装配起来，将移动着的 DNA 聚合酶分子保持在 DNA 上的概括图解。这里显示的简化反应中，夹子装载器一旦完成组装后就解离到溶液中。在真正的复制叉上，夹子装载器与后随链聚合酶保持靠拢，准备好在每个新冈崎片段开始合成的位置组装新夹子（图 5-22）[(A)和(B)摘自 X. -P. Kong et al.,*Cell* 69：425～437，1992. © Elsevier]

在前导链模板上，移动的 DNA 聚合酶紧密地结合在夹子上，且两者的结合能维持相当一段时间。然而，在后随链模板上，每次聚合酶到达前面冈崎片段的 5′端时，就会释放出来；然后，该聚合酶分子再与装配在下一个冈崎片段的 RNA 引物上的另一个夹子结合（图 5-20）。

复制叉上的蛋白质协同作用，形成一个复制机器

尽管我们已经讨论了 DNA 复制，看起来 DNA 复制似乎是多种蛋白质独自发挥作用完成的，实际上，这些蛋白质大部分都形成一个巨大的多酶复合体，它沿着 DNA 迅速移动。该复合体就像一个微型的缝纫机，各种蛋白质构成了它的元件，而核苷三磷酸水解则为其提供了动力。虽然研究复制复合体都集中在大肠杆菌及其他的一些病毒内进行，但是下面我们会看到，在真核生物中也存在非常类似的复合体。

图 5-21 总结了复制机器各个亚单位的功能。有两个 DNA 聚合酶分子在复制叉位置发挥作用，一个位于前导链上，另一个位于后随链上。夹在后随链上的 DNA 聚合酶分子，与在它前面沿着链移动的一个或多个 DNA 解旋酶分子协同作用，打开 DNA 螺旋，协同性地结合的单链 DNA 结合蛋白分子也为螺旋的打开提供帮助。虽然位于前导链上的 DNA 聚合酶分子

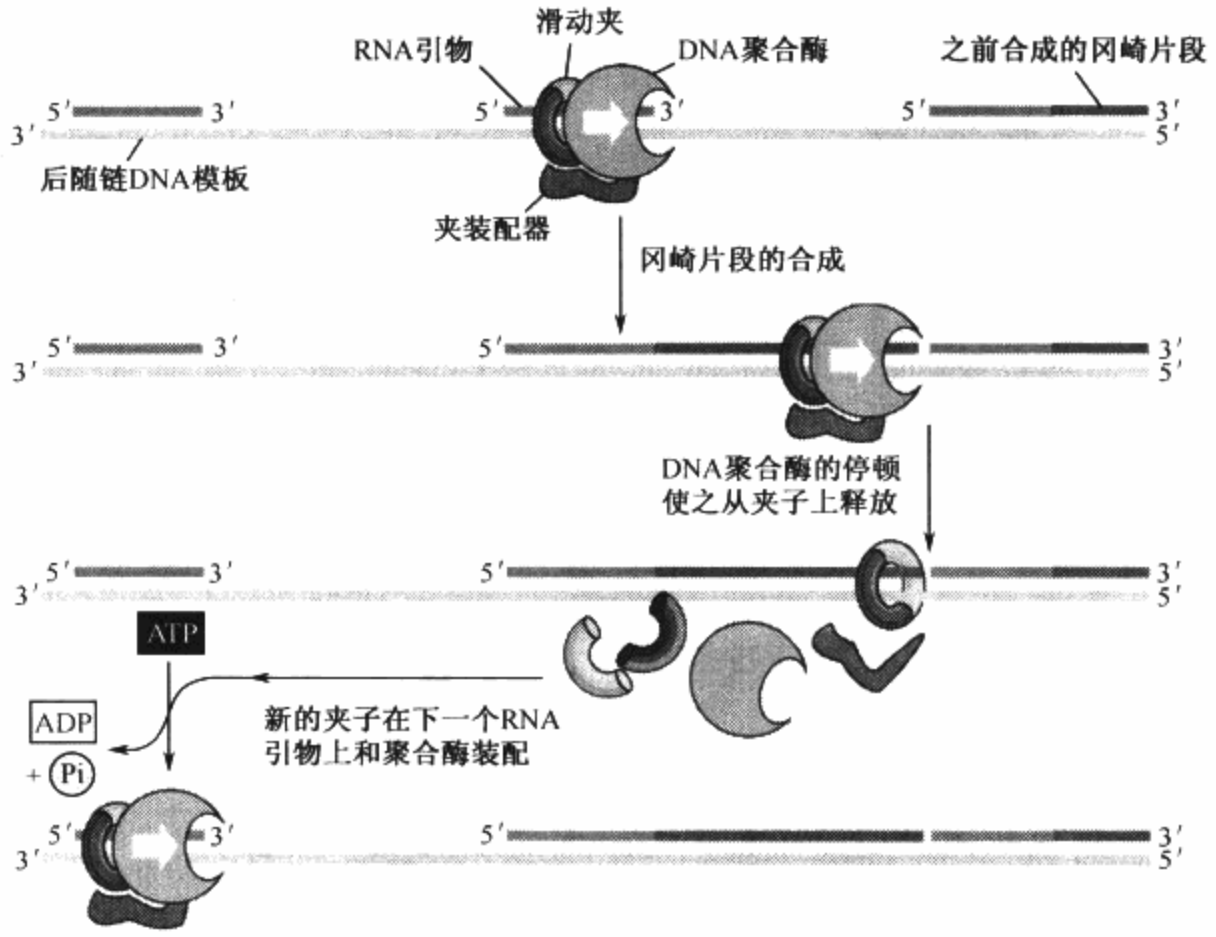


图 5-20　DNA 聚合酶和夹子蛋白质在后随链上的上载和下载循环

为了达到说明的目的，这里显示了夹子装载器和后随链聚合酶的联合；实际上，夹子装载器与复制叉一起，由一个包括前导链和后随链 DNA 聚合酶的复合体携带（图 5-22）

能够以连续的方式发挥作用，但是后随链上的聚合酶分子必须短暂间歇以后重新开始，使用 DNA 引发酶分子合成的一段短 RNA 引物分子作为引物。

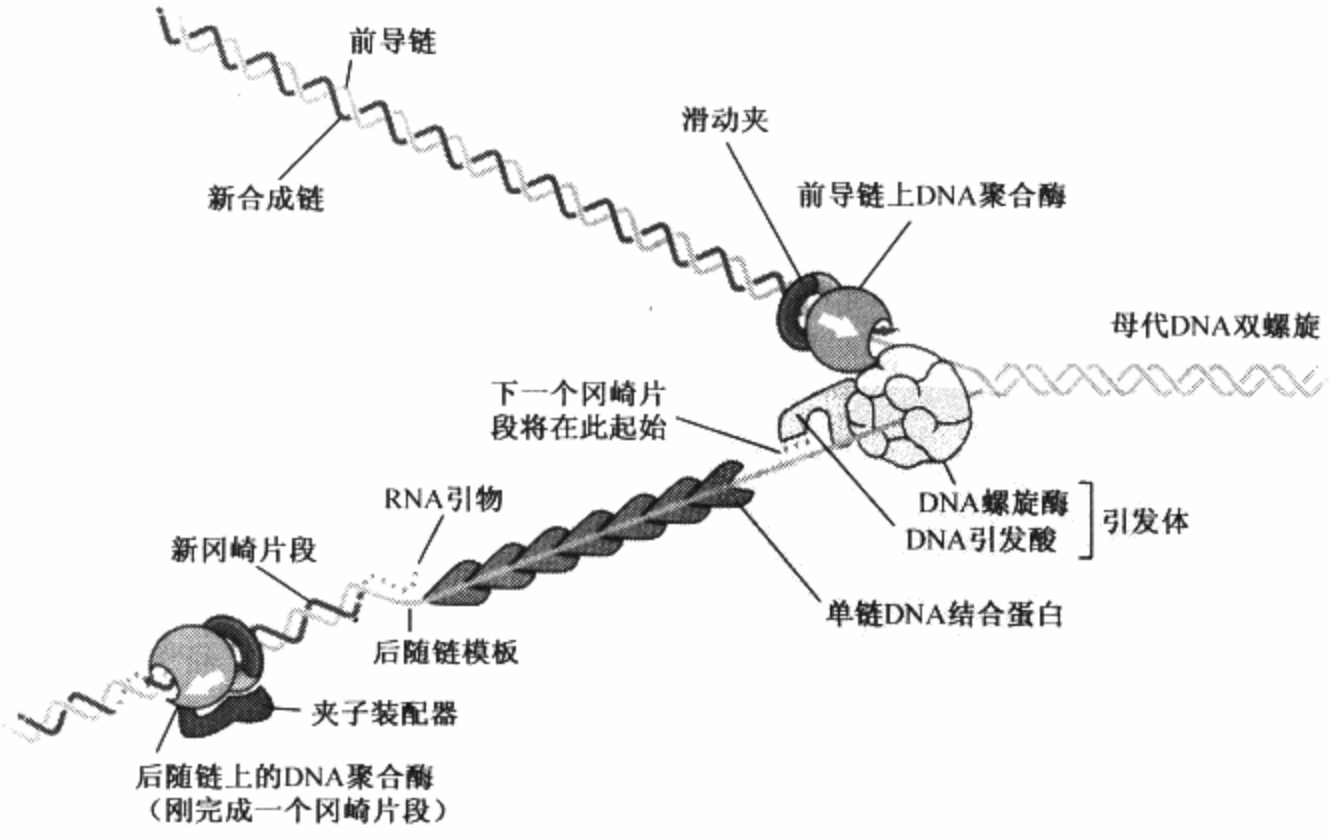
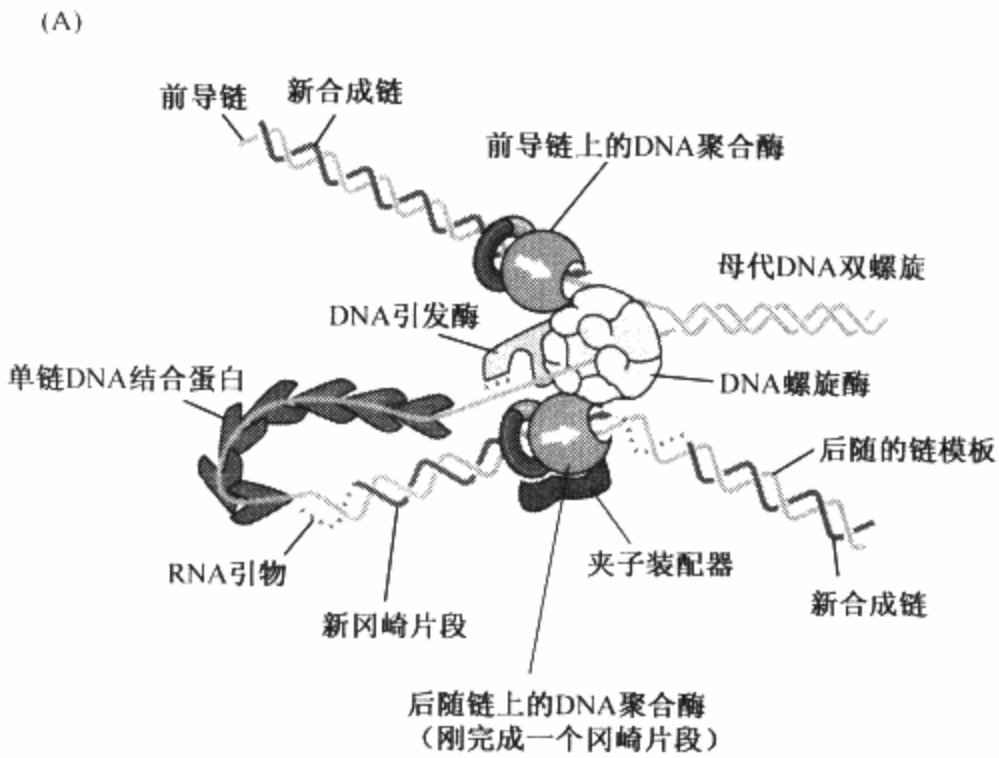
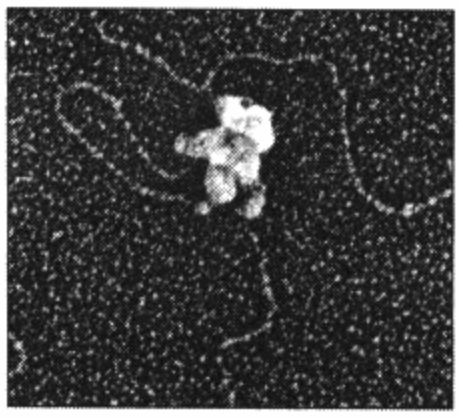


图 5-21　细菌 DNA 复制叉上的蛋白质

图示了在 DNA 复制叉发挥作用的主要类型的蛋白质，显示了它们在 DNA 上大致的位置

所有这些蛋白质组分的紧密结合大大地提高了复制的效率。在原核生物中，引发酶直接与 DNA 解旋酶相连，在后随链上形成一个单位，称为**引发体**（primosome）。引发体由 DNA 解旋酶提供动力，随复制叉一起移动，边走边合成 RNA 引物。同样在后随链上合成 DNA 的 DNA 聚合酶分子与其蛋白质协同移动，合成连串的新的冈崎片段。为了能够适应这些蛋白质的这种排列形式，后随链似乎以图 5-22 中所示的方式回折起来。这种排列形式也促进了每次合成冈崎片段时聚合酶夹子的装载：始终使夹子装载器和后随链 DNA 聚合酶分子处于适当





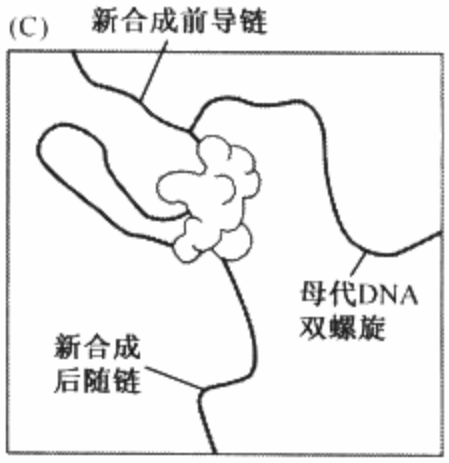


图 5-22　移动的复制叉

（A）该简图显示了目前有关复制叉移动时位于复制叉位置的复制蛋白质排序的观点。图 5-21 中的图稍作改动，将后随链上的 DNA 折叠起来，以引入后随链 DNA 聚合酶和前导链 DNA 聚合酶形成复合物，这个折叠过程也将每个完成的冈崎片段带到下一个冈崎片段合成起始位点附近（与图 5-21 比较）。因为后随链 DNA 聚合酶保持与剩余复制蛋白质结合，所以它可以反复利用合成连续的冈崎片段。在此图中，后随链 DNA 聚合酶即将要释放复制完成的 DNA 片段，并移动到将在附近合成的 RNA 引物上，这是起始下一个 DNA 片段所必需的。注意该图中一条子 DNA 螺旋伸向右下方，另一条则伸向左上方。还有其他的蛋白质帮助将复制叉不同的蛋白质成分维系起来，使它们能够作为一个精细协同的蛋白质装置发挥功能。实际的蛋白质复合体比图示的要更加紧密，夹子装载器由一种此处未显示的相互作用维系起来。（B）显示当噬菌体 T4 沿着模板移动并合成其后面的 DNA 时其复制机器的电镜照片。（C）该简图给出显微图片的解释：特别要注意后随链上的 DNA 突环。显然，在制备该样品用于电镜观察时，复制蛋白质与最前面的复制叉部分分开[(B)经 Jack Griffith 允许使用]

的位置，作为蛋白质装置的一部分，即使它们已从 DNA 上分离。因此，这些参与复制的蛋白质彼此连接起来，形成一个单一的大单位（总分子质量＞10^6Da），沿 DNA 迅速移动，使得 DNA 以一种协同有效的方式在复制叉的两边合成。

在后随链上，DNA 复制机器后面留下大量未封口的冈崎片段，它们仍然含有自 5′端引导合成的 RNA 引物。这个 RNA 被切除产生的缺口由在复制叉后面发挥作用的 DNA 修复酶进行填补（图 5-13）。

一种链指导的错配修复系统清除了逃过复制机器检查的复制错误

前面提到过，像大肠杆菌这样的细菌能够每隔 30min 分裂一次，因而容易对大量的群体进行筛选，找到某一特殊生命过程发生改变的稀有突变的细胞。有一类令人感兴趣的突变体在所谓的增变基因中含有突变，增变基因一旦失活，就会大大增加自发突变的概率。毫不奇怪，这种突变体产生了存在缺陷的 3′→5′校对外切核酸酶，该外切核酸酶是 DNA 聚合酶活性的一部分（图 5-9 和图 5-10）。如果外切核酸酶活性存在缺陷，DNA 就不能再有效地校对了，本来可以清除的许多复制错误在这种情况下就会在 DNA 中积累起来。

对其他表现出异常高的突变率的大肠杆菌突变体进行研究后发现，存在其他的校对系统清除被校对外切核酸酶漏掉的聚合酶产生的复制错误。该**链指导的错配修复**（strand-directed mismatch repair）系统能发现因非互补碱基对之间的错配导致在 DNA 螺旋中产生形变的潜在可能性。但是，如果这种校对系统仅仅识别新复制的 DNA 中的一个错配，并随机校正错配的 2 个核苷酸中的 1 个，那么它就会有一半的概率错误地将原来的模板链进行“校正”，使其与错误的那条配对，因而最终无法从整体上降低错误率。为了能有效地校对，该校对系统必须能够辨别并消除发生复制错误的新合成链上存在的错配核苷酸。

大肠杆菌中错配校对系统利用的链识别机制依赖于在 DNA 中选定的 A 残基上进行甲基化。GATC 序列中所有的 A 残基都增加一个甲基基团，但在 A 整合到新合成的 DNA 链后不久便不会发生甲基化了，其结果就是只有紧跟在复制叉后面的新链中的 GATC 序列还未被甲基化。对这些未甲基化的 GATC 的识别，使得修复机制能瞬间将新链和旧链区别开。如果它们的错配需要选择性的加以清除，这正好满足要求。这个由三步组成的过程涉及错配的识别，含有错配的 DNA 片段从新合成的链上切除，以及利用旧模板对切除片段的重新合成，从而最终移走错配。该链指导的错配修复系统将 DNA 复制过程中产生的错误数量又降低了 2 个数量级（表 5-1）。

类似的错配校对系统也在人类细胞中发挥作用。如果一个人继承了存在缺陷的错配修复基因的一个拷贝（在另一条染色体上存在功能正常的基因），那么他就会有明显的患有某种癌症的倾向，这个事实表明了该系统的重要性。在一个称为遗传性非息肉结肠癌（HNPCC）的结肠癌中，剩余的功能正常的基因自发突变，产生许多体细胞的克隆；由于这些细胞缺乏错配校对机制，因而能够异常迅速地累计突变。大部分癌症都是细胞内积累多种突变后的结果（讨论见第 23 章），因此缺乏错配校对的细胞的癌变的可能性大大增加。幸运的是，我们大多数都继承了每个编码错配校对蛋白的基因的两个完好拷贝；这样就能够保护我们，因为同一细胞中两份拷贝都发生突变的可能性极小。

在真核生物中，在发生错配的位点区分新合成的链与亲代模板链的机制并不依赖于 DNA 的甲基化。事实上，有些真核生物，包括酵母和果蝇的 DNA 都不甲基化。新合成的 DNA 链上优先产生切口，生物化学实验表明，这样的切口（也称为单链断裂）在真核细胞内提供了指导错配校对系统到达合适链的信号（图 5-23）。

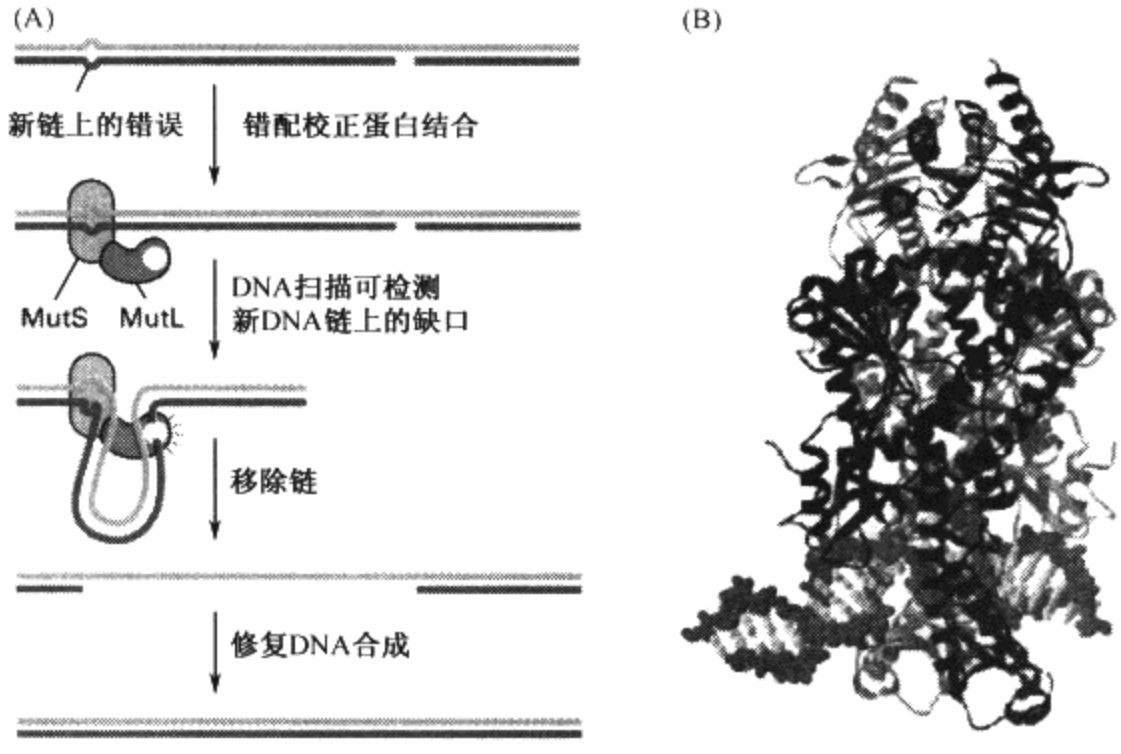


图 5-23　真核生物中链指导的错配修复模型

（A）两种蛋白质在细菌和真核细胞中都存在：MutS 特异结合错配的碱基对，而 MutL 扫描附近的 DNA 以寻找切口。一旦找到一个切口，MutL 引发切口链的降解，一直向错配方向降解直到到达错配位置。因为在真核生物中切口大都限于出现在新复制的链上，所以选择性地将复制错误排除。在细菌中，机制相同，除了复合物中另外一种蛋白质（MutH）将未甲基化的（因而是新合成的）GATC 序列切开，从而在此处开始这里所示的过程。（B）与 DNA 错配结合的 MutS 蛋白的结构。该蛋白质是二聚体，如图所示，它将 DNA 双螺旋紧紧抓住，在错配碱基对位置使 DNA 扭结。似乎 MutS 蛋白通过检验易于扭结的位点来扫描 DNA 寻找错配，这些位点是那些没有正常互补碱基对的位点[(B)摘自 G. Obmolova et al., *Nature* 407：703～710，2000. © Macmilian Magazimes Ltd]

DNA 拓扑异构酶防止复制期间 DNA 乱成团

当复制叉沿双链 DNA 移动时，它会造成所谓的“缠绕问题”。在复制叉位置复制的碱基对中每 10 个对应于亲代双螺旋直径相当的完整一圈。因此，为了让复制叉向前移动，位于复制叉前面的全部染色体通常将不得不迅速旋转（图 5-24）。对于长长的染色体来讲，这将需要大量的能量，因此，细胞使用另一种策略而代之：由一种称为 **DNA 拓扑异构酶**（DNA topoisomerase）的蛋白质在 DNA 螺旋中形成旋转。

DNA 拓扑异构酶可看成是一种可逆的核酸酶，能将自身共价结合到 DNA 骨架的一个磷酸上，从而打断 DNA 链中的一个磷酸二酯键。该反应是可逆的，当拓扑异构酶离开时该磷酸二酯键重新形成。

一种称为拓扑异构酶 I 的拓扑酶，产生短暂的单链断裂（或切口）；这种在磷酸二酯

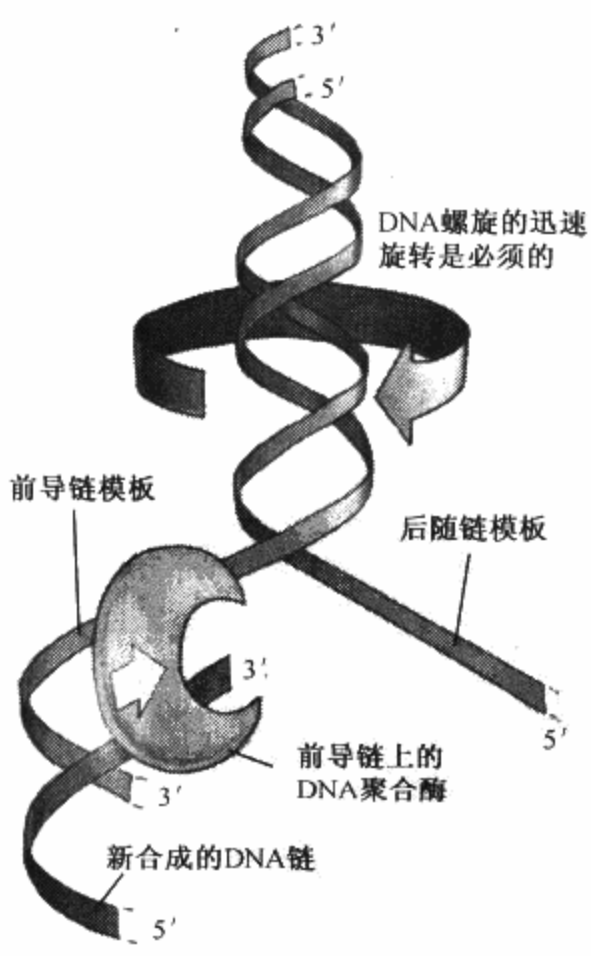


图 5-24　DNA 复制过程中出现的“缠绕难题”

对于一个以每秒 500 个核苷酸的速度移动的细菌复制叉而言，复制叉前面的亲代 DNA 螺旋必须以 50 圈的速度旋转

骨架中的断裂，使得切口边 DNA 螺旋的两部分都能利用与切口相对的链中的磷酸二酯键作为旋转轴，彼此相对做自由旋转（图 5-25）。DNA 螺旋中产生的任何张力都将驱动这种旋转向释放这种张力的方向进行。因此，在只有一小段螺旋旋转的情况下 DNA 复制只发生在复制叉前面的那部分。DNA 转录（讨论见第 6 章）时出现的类似缠绕问题也可以相同的方式得以解决。因为连接 DNA 拓扑异构酶和 DNA 磷酸之间的共价键保留了断裂的磷酸二酯键的能量，重新连接是迅速的，且并不需要另外的能量输入。从这个角度看，再连接的机制与前面讨论的 DNA 连接酶催化的连接反应是不同的（图 5-14）。

另一种拓扑异构酶，拓扑异构酶Ⅱ，同时与 DNA 螺旋的两条链共价连接起来，在螺旋中形成短暂的双链断裂。这些酶由染色体上两条双螺旋彼此发生交叉的位点激活。一旦拓扑异构酶Ⅱ分子结合到这种交叉位点，该蛋白质就利用 ATP 水解释放的能量有效地进行下面的一系列反应：①它将其中一条双螺旋 DNA 可逆地打断，形成一个 DNA“入口”；②它使附近的另一条双螺旋 DNA 通过这个断口；③然后它重新封闭这个断口，并从 DNA 上解离（图 5-26）。通过这种方式，Ⅱ型 DNA 拓扑异构酶能有效地将互相连锁的 DNA 环分开(图 5-27)。

同样的反应也防止产生严重的 DNA 缠绕成团问题，否则在 DNA 复制过程中就会出现这种难题。这种作用可以用酵母的某些突变细胞来很好地解释，这些酵母的突变细胞能产生一种在 37℃失活的拓扑异构酶Ⅱ，而不是正常的拓扑异构酶。当突变细胞加热到这个温度时，它们的子代染色体完成复制后相互缠绕在一起，不能分开。对于任何想要费力地解开钓鱼线上的结，而又不使用剪刀的人来说，很容易体会到拓扑异构酶Ⅱ对于解开染色体缠绕的巨大用处。

真核生物和细菌的 DNA 复制过程相似

我们对 DNA 复制的了解大多数最初都来自于体外能进行 DNA 复制的细菌和噬菌体纯化的多酶系统的研究。19 世纪 70 年代，由于对多种复制基因突变体的优先分离，大大促进了这些系统的发展；这些突变体被用于鉴定和纯化相应的复制蛋白质。19 世纪 80 年代中期，有人描述了第一个在体外精确复制 DNA 的哺乳动物复制系统。到目前，在酿酒酵母中已分离了几乎所有复制组分的编码基因的突变体。因此，已经知道大量的真核生物 DNA 复制的详细知识，而且很明显，DNA 复制的基本特征（包括复制叉几何结构以及利用多蛋白质复制机器）在将细菌和真核生物分开的长期进化过程中，一直是保守的。

真核生物复制机器中的蛋白质组分比原核生物的要多，尽管它们的基本功能一样。因此，例如，真核生物的单链结合蛋白（SSB）由 3 个亚基组成，而细菌中只有 1 个亚基。同样，DNA 引发酶也整合到一个多亚基酶中，称其为 DNA 聚合酶 α。聚合酶 α 以 RNA 为引物在后随链上起始每一个冈崎片段的合成，然后，在将引物 3′端传递给另一种酶 DNA 聚合酶 δ 之前，将 RNA 引物延伸，合成一小段 DNA，然后 DNA 聚合酶 δ 在夹子蛋白的帮助下，合成冈崎片段的剩余部分（图 5-28）。

在下一部分中我们会看到，真核生物复制机器面临额外的难题，因为它必须通过核小体进行复制，核小体是第 4 章讨论的染色体重复的结构单位。沿 DNA 上每隔 200 个核苷酸对存在 1 个核小体，这也许可以理解为什么真核生物中，后随链上，每隔 100～200 个核苷酸对合成新的冈崎片段，而不是像细菌中那样隔 1000～2000 个核苷酸对合成一条。核小体也可能作为一种障碍物起作用，使 DNA 聚合酶分子的移动减速。这可能是为什么真核生物复制叉的移动速度仅为细菌复制叉的 1/10 的原因。

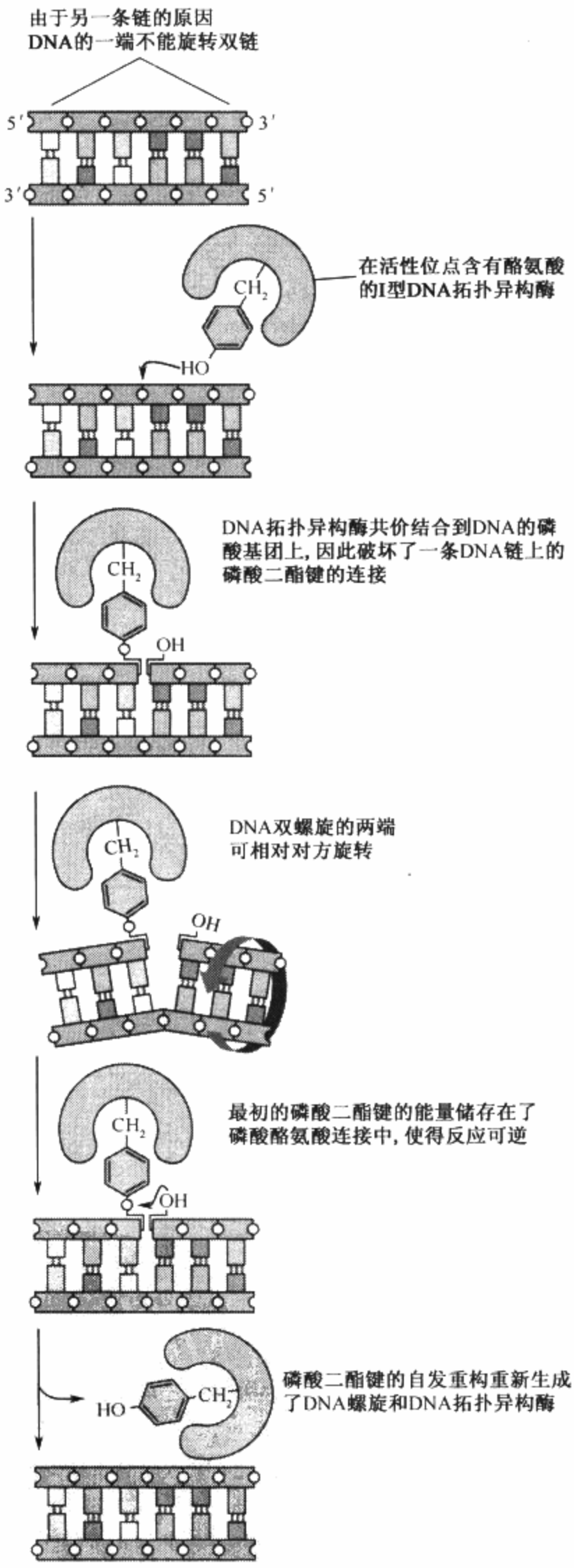


图 5-25　一种真核 DNA 拓扑异构酶 I 催化的可逆切割反应

如图所示，这些酶瞬间与 DNA 形成一个共价键；这使得 DNA 能绕连接有蓝色磷酸的共价骨架链自由旋转

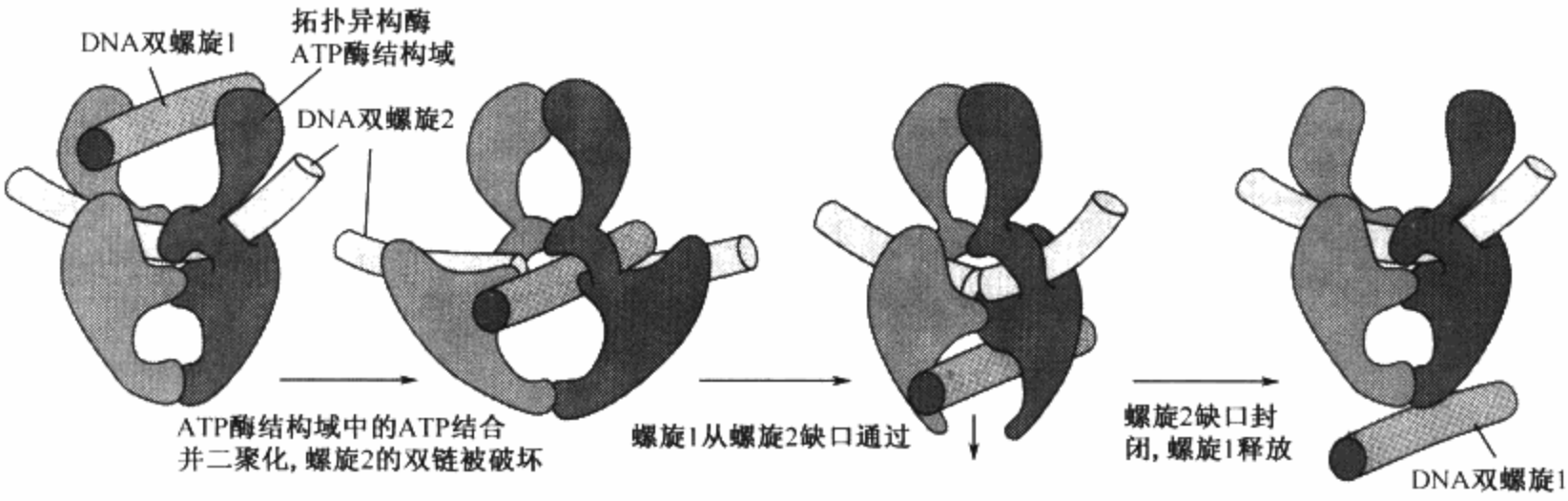


图 5-26　拓扑异构酶Ⅱ作用的模型

如图所示，ATP 与两个 ATP 酶结构域的结合导致它们二聚化，并驱动所示的反应。因为存在不可水解的 ATP 类似物时这个反应可以进行一个循环，所以认为 ATP 水解只需用来使酶复位，以便催化每一轮新反应。这个模型基于对该酶的结构和机制研究（改编后摘自 J. M. Berger，*Curr. Opin. Struct. Biol.* 8：26～32，1998）

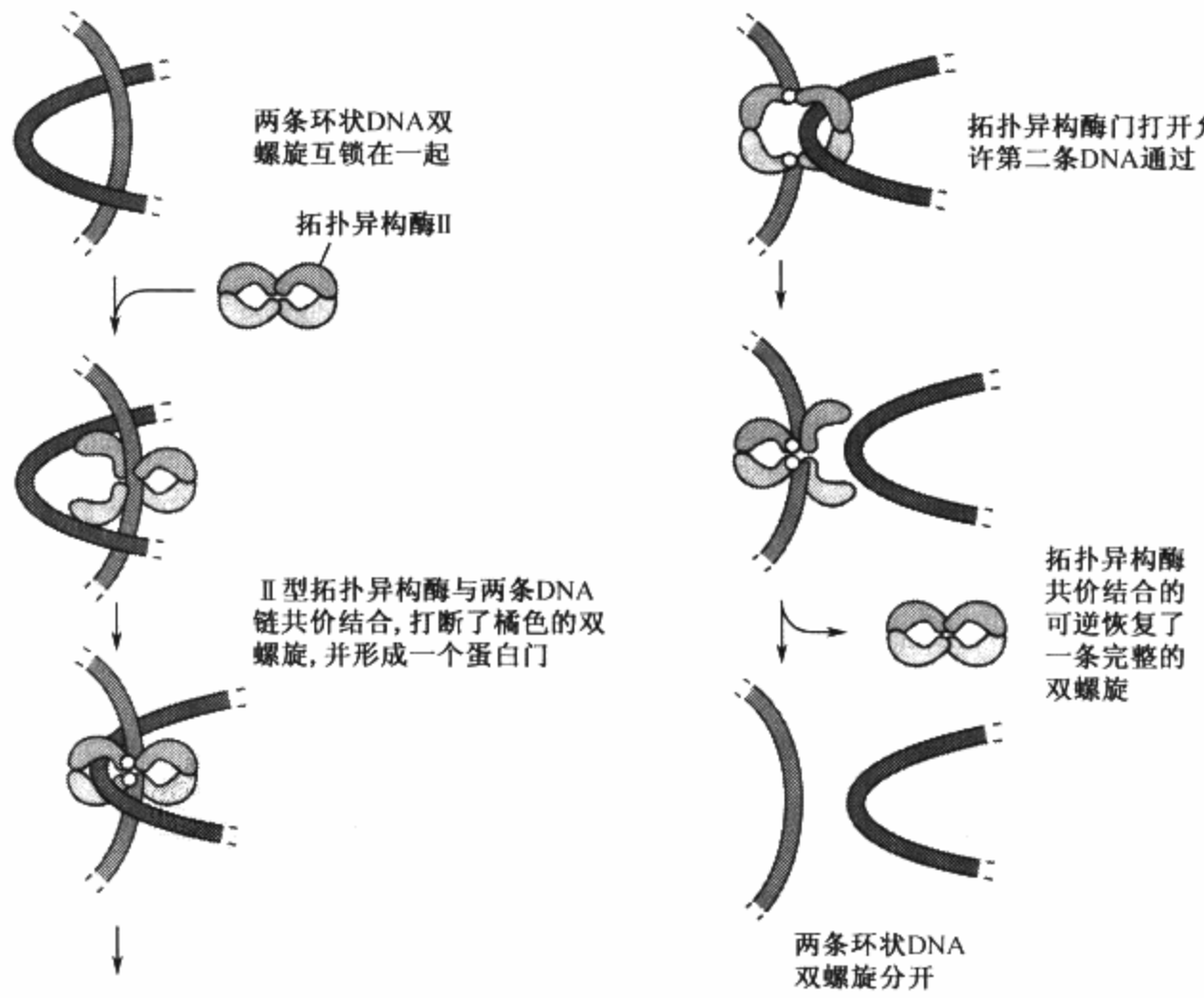


图 5-27　DNA 拓扑异构酶Ⅱ催化的 DNA 螺旋通过的反应

细胞内利用相同的反应解开 DNA。与拓扑异构酶Ⅰ不同，拓扑异构酶Ⅱ使用 ATP 水解作用，有些细菌中的拓扑异构酶 II 能够向 DNA 中引入超螺旋。拓扑异构酶Ⅱ大都局限于在真核生物增殖的细胞中发挥作用；作为部分原因，它们一直是抗癌药物常用靶位点

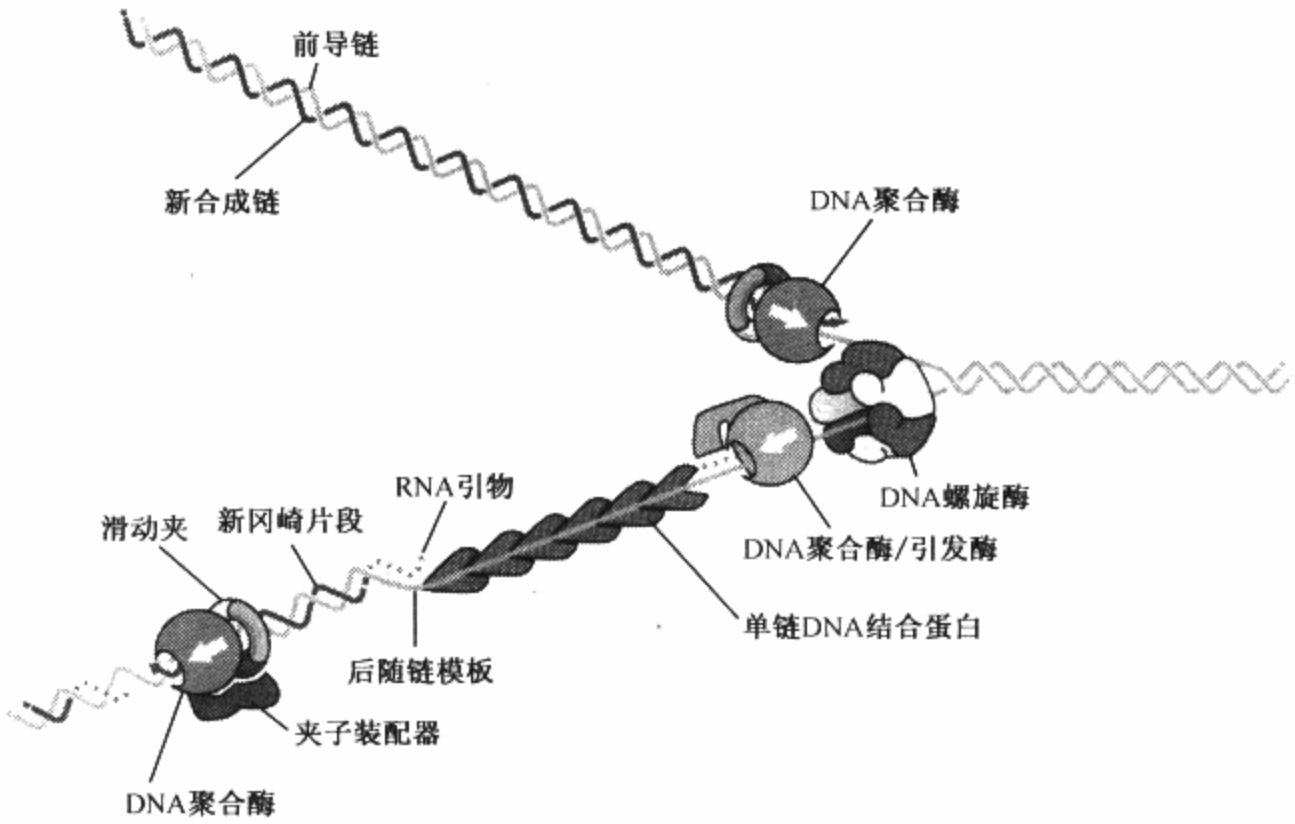


图 5-28　哺乳动物的复制叉

绘制此复制叉的目的是为了强调与图 5-21 中描绘的细菌复制叉的相似性。尽管两个复制叉利用相同的基本组分，但哺乳动物复制叉至少在两个重要的方面存在差异。第一，它在后随链上利用两种不同的 DNA 聚合酶。第二，哺乳动物 DNA 引发酶是其中一个后随链 DNA 聚合酶——DNA 聚合酶 α 的亚基之一，而细菌的引发酶与引发体中的一个 DNA 解旋酶结合。聚合酶 α（及与其结合的引发酶）开始时合成 RNA，然后从它延伸 DNA，接着把 DNA 链交给的第二个聚合酶（δ），由它进行延伸。还不清楚为什么真核生物 DNA 复制在后随链上需要两种不同的聚合酶。哺乳动物的主要的 DNA 解旋酶似乎以 6 种不同的 Mcm 蛋白形成的环为基础；该环可以沿前导链移动，而不是沿这里所示的后随链模板移动

小结

DNA 复制在称为复制叉的 Y 形结构处进行。一种能自我校正的 DNA 聚合酶催化核苷酸由 5′→3′方向聚合，以高度的忠实性对一条 DNA 模板链进行拷贝。由于 DNA 双螺旋的两条链是反向平行的，这种 5′→3′方向的合成只能在复制叉处的一条链（前导链）上持续进行。在后随链上，短的 DNA 片段必须通过“倒缝”过程合成。因为自我校正 DNA 聚合酶不能起始新链的合成，所以这些后随链的 DNA 片段由短 RNA 引物分子引发，随后再切除这些 RNA 分子并用 DNA 替换。

DNA 复制需要许多蛋白质的协同作用。这些蛋白质包括：①用于催化核苷三磷酸聚合的 DNA 聚合酶和 DNA 引发酶；②用于帮助打开 DNA 螺旋，以便能够对其进行拷贝的 DNA 解旋酶和单链 DNA 结合（SSB）蛋白；③用于连接不连续合成的后随链 DNA 片段的 DNA 连接酶和一种降解 RNA 引物的酶；④帮助释放螺旋缠绕和 DNA 缠绕成团问题的 DNA 拓扑异构酶。这些蛋白质中有许多在复制叉位置彼此结合起来，形成一种高效的“复制机器”，通过该装置，对单个组分的活性和空间运动进行协调。

染色体上 DNA 复制的起始与完成

我们已经看到一系列的复制蛋白质如何快速并精确地在移动着的复制叉后面合成出两条子代 DNA 双螺旋。但是，首先这个复制机器是怎样组装起来的呢？复制叉又是如何在双链 DNA 分子上形成的呢？在这一部分，我们讨论 DNA 复制是如何起始的，以及细胞又是如何仔细调节这个过程的，以确保起始发生在染色体适当的位置和细胞生命适当的时期。我们还讨论真核生物细胞中复制机器必须克服的特殊难题。这些难题包括需要复制真核染色体中巨大的 DNA 长分子，以及对核小体内与组蛋白紧密地形成复杂结构的 DNA 分子进行拷贝。

DNA 复制起始于复制起点

前面讨论过，DNA 双螺旋通常很稳定：两条 DNA 链通过在各自所含碱基之间形成大量的氢键，牢牢地闭锁在一起。为了能作为模板使用，必须首先打开双螺旋，将两条链分开，暴露出未配对的碱基。我们将会看到，DNA 复制过程开始于特殊的起始蛋白结合到双链 DNA 上并将两条链撬开，打开碱基间的氢键。

DNA 双螺旋上首先被打开的位置称为**复制起点**（replication origin）（图 5-29）。在简单细胞，如细菌或酵母的细胞中，复制起点包括长达几百个核苷酸对的 DNA 序列。这段 DNA 含有吸引起始蛋白的短序列以及特别容易打开的连续 DNA 片段。我们从图 4-4 中看到，A-T 碱基对比 G-C 碱基对之间的氢键数目少。因此，富含 A-T 碱基对的 DNA 分开相对容易，且富含 A-T 碱基对的 DNA 区域经常位于复制起始位置。

尽管图 5-29 描绘的复制叉起始的基本过程对于细菌和真核生物来说是相同的，但是在这两种生物体中，这个过程完成和调控的详细方式却不相同。我们首先看看简单一些且了解更清楚的细菌中复制起始的过程，然后再讨论酵母、哺乳动物和其他真核生物中更复杂的情况。

细菌染色体只有一个 DNA 复制起点

大肠杆菌的基因组位于一个含 4.6×10^6 个核苷酸对的环状 DNA 分子中。DNA 的复制起始于单一的一个复制起点，在这里形成两个复制叉，以相反的方向前进（500～1000 个核苷酸/s），直到它们在染色体中间位置附近相遇（图 5-30）。大肠杆菌能控制 DNA 复制的唯一位点是起始：一旦在起点形成复制叉，它们会以相对稳定的速度移动，直到复制完成。因此，毫不奇怪 DNA 复制的起始是一个高度受调控的过程。起始开始于起始蛋白以多个拷贝结合到复制起点中的特定位点，将 DNA 缠绕在蛋白质周围形成一个巨大的蛋白质-DNA 复合物。然后该复合物结合上一个 DNA 解旋酶，将其放置在邻近的 DNA 单链上，DNA 单链上的碱基因起始蛋白-DNA 复合物的组装而暴露出来。DNA 引发酶结合到解旋酶上，形成引发体，然后它从起点向前移动，合成出一条 RNA 引物，起始第一条 DNA 链的延伸（图 5-31）。这个过程迅速导致剩余蛋白质组装起来，与自起点反方向移动的蛋白质复合物一起，形成两个复制叉。这些蛋白质机器持续合成 DNA，直到每个复制叉下游的每个 DNA 模板都得到复制。

在大肠杆菌中，起始蛋白与复制起点的相互作用受到仔细的调控，使得起始只在细菌有足够营养物质完成整轮复制的时候才会发生。不仅起始蛋白的活性受到调控，而且刚刚使用过的复制起点遇上“不应期”，该不应期是由于新合成的 A 核苷酸的甲基化延迟引起的。复制的再次起始受到了阻碍，直到这些 As 都被甲基化为止（图 5-32）。

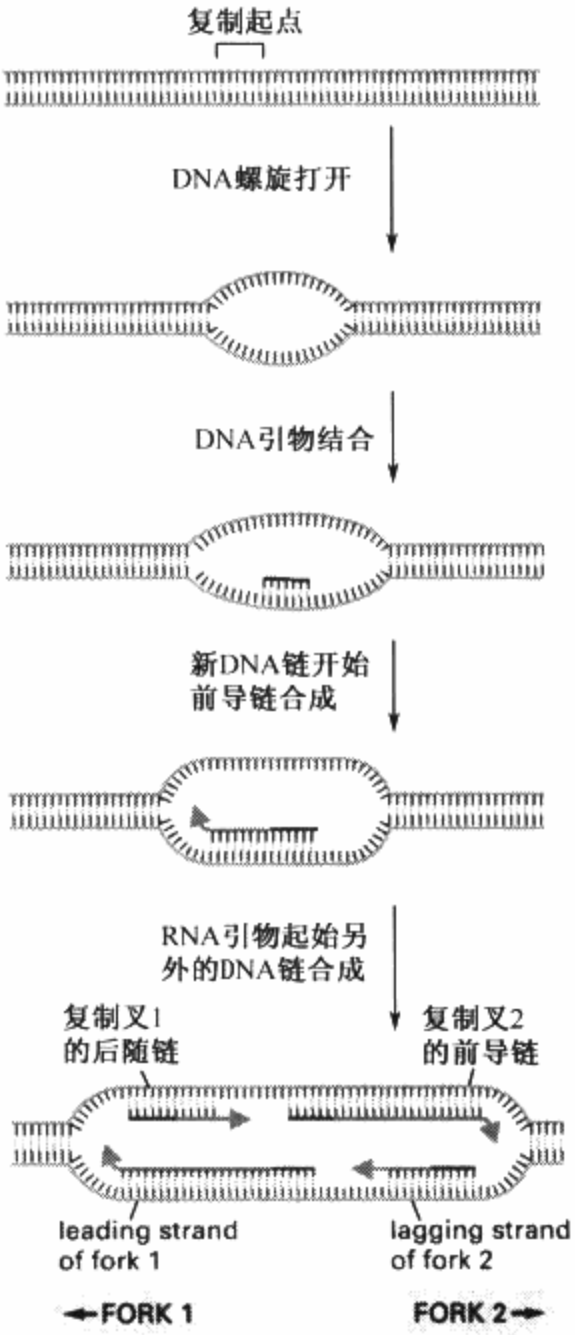


图 5-29　复制叉的起始所形成的复制泡

该图概括了在复制起点参与复制叉的起始的主要步骤。在最后一步形成的结构，称作复制泡，其中亲本 DNA 两条链都彼此分离并作为 DNA 合成的模板

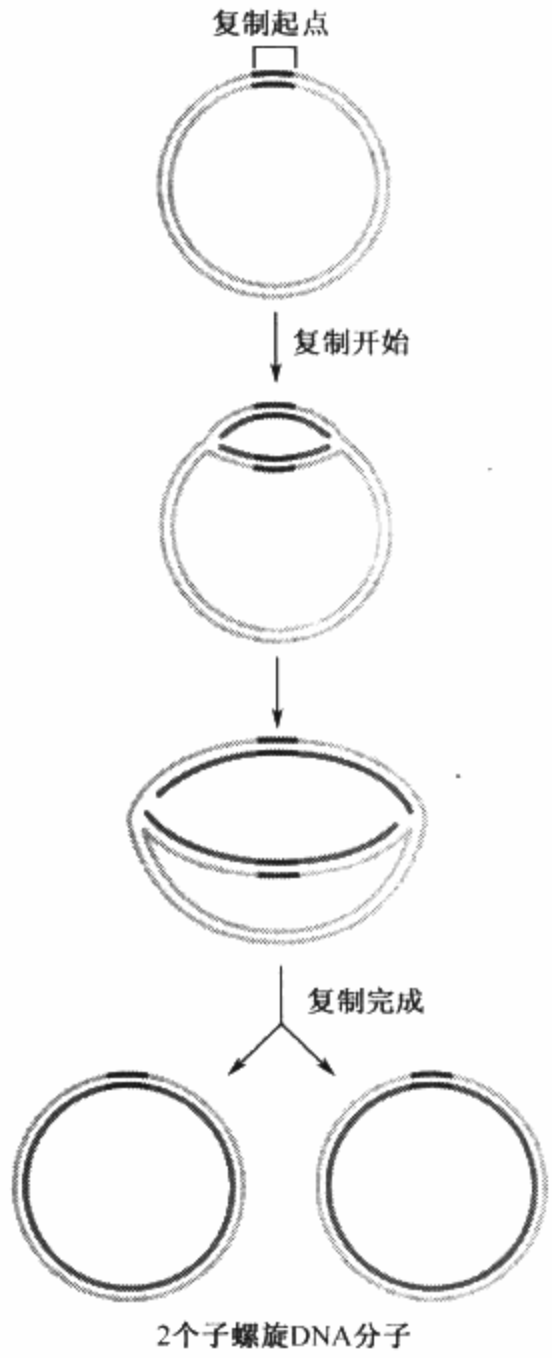


图 5-30　细菌基因组的 DNA 复制

大肠杆菌复制其 4.6×10^6 核苷酸对的基因组需要 40min。为了简便起见，在后随链上未显示冈崎片段。当这两个复制叉彼此靠近并在一轮复制的末端相遇时发生了什么还不十分清楚，虽然知道作为该过程的一部分，引发体发生解离

真核生物的染色体含有多个复制起点

我们已看到细菌中两个复制叉如何从一个复制起点开始，以相反的方向前进，离开起点直到仅有的一条环状染色体中的 DNA 全部复制。细菌基因组小到足以让这两种复制叉在 40min 内就能对基因组进行拷贝。因为大多数真核生物的染色体要大得多，所以需要一种不同的策略，以便适时地进行复制。

19 世纪 60 年代建立了一种方法，用于确定染色体复制的一般模式。用 ^{3}H-胸腺嘧啶对培养基中生长的人类细胞进行短时间标记，这样在这段时间内合成的 DNA 就会具有很高的放射性。然后温和地裂解细胞，将 DNA 划在感光乳剂包被的载玻片表面。通过一种称为放射

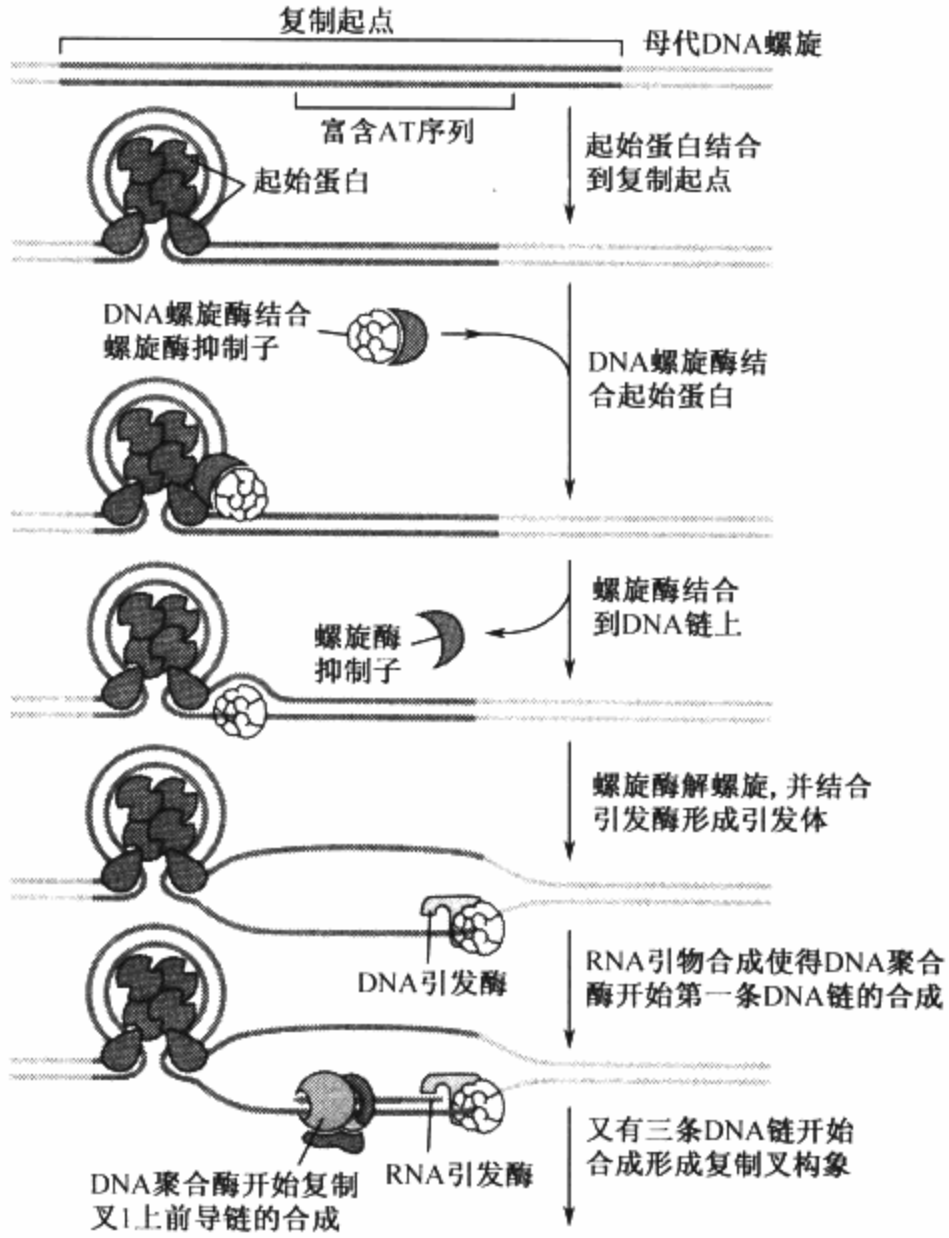


图 5-31　细菌中起始 DNA 复制的蛋白质

图中所示的机制是通过体外研究高度纯化的蛋白质混合物之后建立起来的。对于大肠杆菌的 DNA 复制而言，主要的起始蛋白质是 dnaA 蛋白；引发体由 dnaB（DNA 解旋酶）和 danG（DNA 引发酶）蛋白组成。在溶液中，解旋酶受一种抑制蛋白质（dnaC 蛋白）束缚，该蛋白质由起始蛋白质激活，以便将解旋酶装载到复制起点 DNA 上，然后释放。这个抑制蛋白质防止解旋酶错误地进入细菌基因组上的其他单链 DNA 片段。随后的步骤导致另外的三条 DNA 链的起始（图 5-29），其路途的细节尚未详细描述

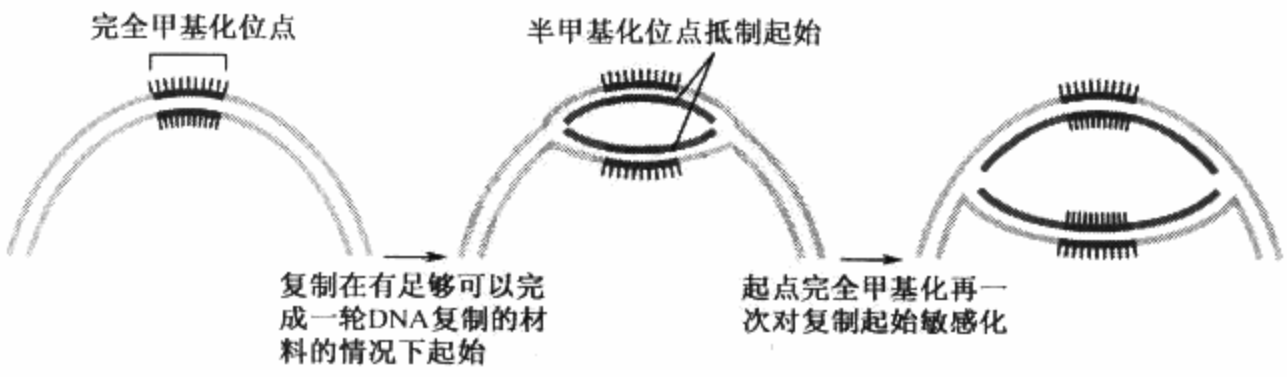


图 5-32　大肠杆菌复制起点的甲基化产生 DNA 起始的不应期

DNA 甲基化发生在 GATC 序列上，它们全都位于复制起点（跨越大约 200 个核苷酸对）。复制起始后约 10min，由 DNA 甲基化酶将半甲基化的起点完全甲基化。前面讨论过，大肠杆菌错配校对系统也利用 GATC 序列复制后甲基化的延迟来区别新合成的 DNA 链和亲本 DNA 链；在这种情况下，有关的 GATC 分散在整个基因组不同位置。只有一种酶，dam 甲基化酶，负责大肠杆菌 GATC 序列的甲基化

自显影的技术，感光乳剂的冲洗显示标记的 DNA 的模式。对放射标记分配的时间加以选择，以便每个复制叉沿 DNA 移动几个微米，这样复制得到的 DNA 在光学显微镜下呈现出线状排列的银颗粒，即使 DNA 分子本身太细小，不能观察到。使用这种方法能够确定复制叉移动的速率以及方向（图 5-33）。根据复制的 DNA 足迹长度随标记时间增加而增加的速率，可以估计出复制叉以每秒 50 个核苷酸的速率移动。这个速度大约是细菌复制叉移动速度的 1/10，这可能反映出紧密包装在染色体中的 DNA 复制起来更加困难。

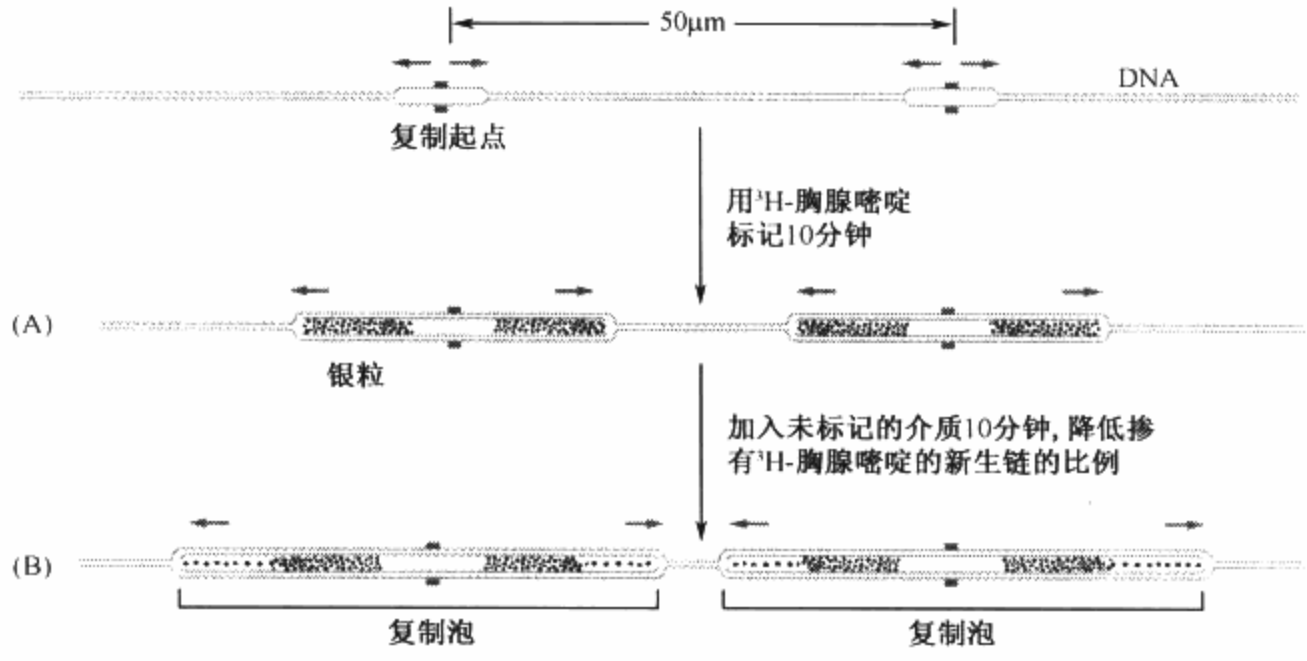


图 5-33　证明真核生物染色体上复制叉的形成和移动的实验

培养的人细胞中新合成的 DNA 用高放射性的胸腺嘧啶（^{3}H-胸腺嘧啶）脉冲短时间标记。（A）在该试验中，将细胞裂解，并在载玻片上将 DNA 展开，接着用感光乳剂覆盖载玻片。几个月以后，将感光乳剂冲洗出来，在放射性 DNA 上显示出一条银颗粒。此图中显示褐色 DNA 只是为了帮助解释放射自显影；在这样的实验中，未标记的 DNA 是看不见的。（B）该实验除进一步在未标记培养基中温育，使其他放射性水平较低的 DNA 能够复制外，其他方面都一样。发现（B）中的这对黑色印迹具有反方向的逐渐消失的银颗粒，表明复制叉从形成复制泡的中间复制起点双向进行移动（图 5-29）。只有当复制叉遇到以反向移动的复制叉或者当它到达染色体末端时，才认为它会停下来；通过这种方式，所有 DNA 最终都得到复制

一条平均大小的人类染色体含有一条长约 1.5×10^8 核苷酸对的线性 DNA 分子。如果只有一个复制叉以 50 个核苷酸/s 的速度移动，从头到尾复制这样一条 DNA 分子，将需要 $0.02\times150\times10^6=3.0\times10^6$ s（大约 800h）。因此，正如所期望的那样，放射自显影实验正好揭示出每条真核染色体上有许多复制叉在同时移动。此外，发现许多复制叉在同一 DNA 区域彼此靠拢，而同一 DNA 其他区域却不存在复制叉。

进一步进行这类实验后发现：①复制起点倾向于以 20～80 个起点成簇激活，这些成簇的复制起点称为复制单位。②新的复制单位在细胞周期中似乎以不同的次数被激活，直到所有的 DNA 都得到复制，这是我们下面将要谈到的问题。③在一个复制单位内，单个复制起点彼此间隔 30 000～300 000 个核苷酸对。④与细菌中一样，复制叉成对产生，并在它们同一起点反向离开时形成一个复制泡，只有当它们迎头碰到反向而行的复制叉时，才会停下来（或者当它们到达染色体末端）。通过这种方式，许多复制叉在每条染色体上能够独自进行复制，但仍然合成两条完整的子代 DNA 双螺旋。

在真核生物中 DNA 复制只在细胞周期的一段时间内进行

细菌在快速生长时，连续复制它们的 DNA，在上一轮 DNA 复制还未完成时，就能起始新的一轮复制。相比之下，真核生物细胞 DNA 的复制只在细胞分裂周期的特定时期进行，称为 DNA 合成期或 **S 期**（S phase）（图 5-34）。在哺乳细胞中，S 期通常持续大约 8h；在更简单一些的真核细胞如酵母中，S 期可以短至 40min。到 S 期结束时，每条染色体都已复制完毕，产生两条完整的拷贝，它们在中心粒位置连接在一起，直到不久后进入 M 期（M 指有丝分裂）。在第 17 章中，我们讨论控制细胞周期运行的系统，并解释为什么每次进入细胞周期时，都要求已经成功地完成了上一期。

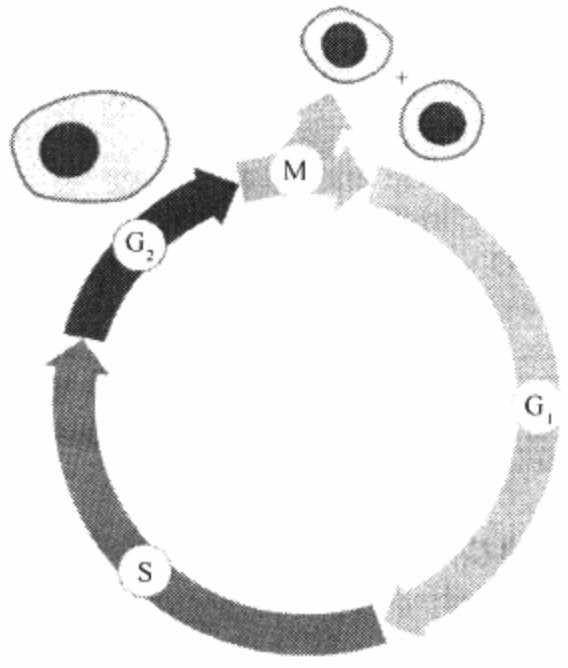


图 5-34　标准的真核生物细胞周期的 4 个连续阶段

在 G_1、S 和 G_2 期，细胞持续进行生长。在 M 期，生长停止，细胞核分裂，细胞分裂成两个子细胞。DNA 复制只限于在期间的某段时间内进行，称为 S 期。G_1 期是 M 期和 S 期之间的间隙；G_2 期是 S 期与 M 期之间的间隙

在下面的内容中，我们探讨在细胞周期的 S 期中染色体的复制是如何协调进行的。

同一染色体上的不同区域在 S 期的不同时间进行复制

在哺乳动物细胞内，如果考虑复制叉的移动速度，以及复制单位中起点之间的最大距离，那么位于相邻复制起点间的区域中的 DNA 完成复制一般只需要大约 1h。然而，哺乳动物细胞 S 期通常持续 8h。这意味着复制起点并不是同时激活，并且每个复制单位中的 DNA（上面我们注释过，它含有 20～80 个成串的复制起点）只在整个 S 期的一小段时间内进行复制。

是不同的复制单位随机激活，还是基因组的不同区域按制定的秩序复制呢？弄清这个问题的一个方法就是利用胸腺嘧啶类似物溴代脱氧尿嘧啶核苷（BrdU），在 S 期的不同时间加入 BrdU，标记经同步化的细胞群中新合成的 DNA。之后，在 M 期，有丝分裂染色体 DNA 中那些整合了 BrdU 的区域就能通过它们改变的染色特性或者抗 BrdU 的抗体加以识别。结果表明，每条染色体的不同区域在 S 期以可重复的秩序进行复制（图 5-35）。此外，正如从

DNA 放射自显影图片中看到的成串的复制叉所推测到的那样（图 5-33），在一大段染色体区域内复制的时间安排得以协调。

高度压缩的染色质复制较晚，而压缩程度较小的染色质中的基因倾向于较早复制

看起来复制起点激活的顺序似乎部分依赖于复制起点所在的染色质结构。在第 4 章中我们看到异染色质是染色质的一种特别压缩的状态，而活跃转录的染色质具有压缩程度较小的构型，RNA 合成明显需要这种构型。异染色质倾向于在 S 期的晚期才复制，这意味着复制的时间安排与染色质中 DNA 的包装有关。对雌性哺乳动物细胞的两条 X 染色体的检验证实了这条推理。虽然这两条 X 染色体含有基本相同的 DNA 序列，但是一条进行活跃的 DNA 转录，而另一条则不进行活跃的 DNA 转录。几乎所有的失活 X 染色体都压缩成异染色质，并在 S 期的晚期进行 DNA 复制。而另一条活跃的 X 染色体则压缩程度较小，并在整个 S 期都进行复制。

这些发现意味着基因组中染色质压缩程度较小，因而复制机器更容易进入的那些区域首先得到复制。放射自显影图像表明在整个 S 期复制叉以类似的速度移动，这样染色体压缩程度似乎就会影响它们的起始时间，而不是它们形成以后移动的速度。

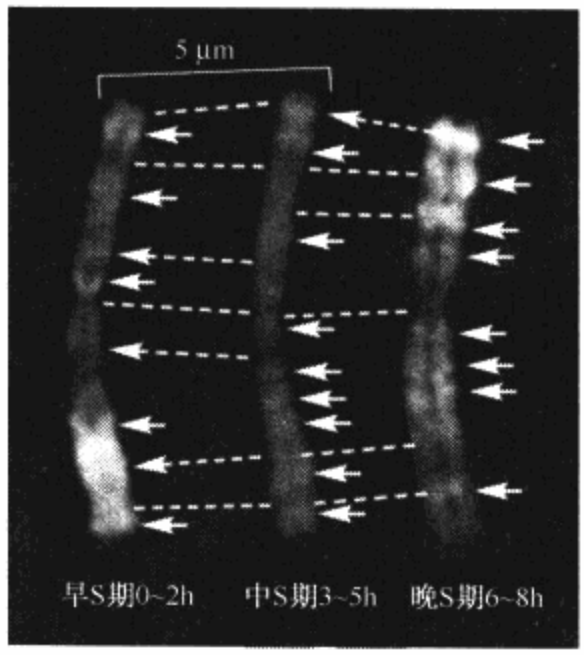


图 5-35　一条染色体的不同区域在 S 期的不同时刻复制

这些光学显微照片显示了染过色的有丝分裂染色体，其中在 S 期之前的不同确定的间隔期内分别对复制的 DNA 进行标记。在这些实验中，细胞首先在存在 BrdU（一种胸腺嘧啶类似物）而不存在胸腺嘧啶时生长，以一致地标记 DNA。然后分别在早、中、晚 S 期不存在 BrdU 时用胸腺嘧啶短时脉冲细胞。因为在胸腺嘧啶脉冲期间合成的 DNA 双螺旋有一条链上是胸腺嘧啶，另一条链上是 BrdU，所以它比剩余 DNA（两条链上都是 BrdU）染色更深，并以一条亮带（箭头）出现在这些底片上。折线将图中所示的 3 个相同染色体拷贝上对应的位置连接起来（经 Elton Stubblefield 允许后使用）

上面染色质结构和 DNA 复制时间安排之间的关系也可以通过测量特定基因的复制次数的研究来证实。结果表明，在所有细胞都活跃表达的所谓“持家基因”，在检测的所有细胞内都在 S 期早期复制。相比之下，只在少部分细胞类型中活跃表达的基因通常在这些细胞内复制时间较早，而在其他类型细胞中复制较晚。

在酿酒酵母中也检验了染色质结构与复制时间安排之间的这种关系。在一种情况下，将一个在 S 期晚期发挥作用并位于染色体转录沉默区域的复制起点，通过实验方法改变位置放到一个转录活跃的区域。换位以后，该复制起点就在 S 早期发挥作用了，这表明在 S 期利用该复制起点的时期是由该复制起点在染色体上的位置决定的。然而，对其他酵母复制起点的研究发现，就是位于常染色质中，也存在一些复制起点在晚期起始复制。因此，复制起点利用的时机可以由它所处的染色质结构以及它的 DNA 序列共同决定。

在一种简单真核生物芽殖酵母中确定的 DNA 序列充当复制起点

真核生物染色体使用许多复制起点进行复制，这些起点中的每个都在细胞周期的 S 期中某一特征性的时间“点火”，讨论过这些以后我们再了解这些复制起点的本质。在本章前面，我们看到细菌中的复制起点有精确的定义，它是指一段特定的 DNA 序列，能使 DNA 复制机

器在 DNA 双螺旋上组装起来，形成一个复制泡，并向相反的方向运动产生复制叉。通过比较，有人可能会推测真核生物染色体中的复制起点也是特定的 DNA 序列。

真核细胞染色体复制起点的搜寻工作在芽殖酵母酿酒酵母中最为成功。利用某个关键基因缺陷的酵母细胞突变体，已经设计了一些强大的筛选方法。这些细胞只有提供了携带该基因功能正常的 DNA，才能在选择培养基上存活。如果将带有该基因的环形质粒直接引入酵母突变体细胞，它将不能进行复制，因为它缺少一个有功能的复制起点。然而，如果将酵母 DNA 的随机片段插入这质粒，就只有少数携带酵母复制起点的质粒 DNA 分子能够进行复制。携带这种质粒的酵母细胞能够进行增殖，是因为以可复制且能遗传给后代细胞的形式提供给它们这个关键基因（图 5-36）。通过其在这些存活的酵母细胞中分离出的质粒中的存在得以鉴定的一段 DNA 序列称为自主复制序列（ARS）。已证实大部分的 ARS 都是真正的复制起点，因此证明获得它们的方法是最有效的。

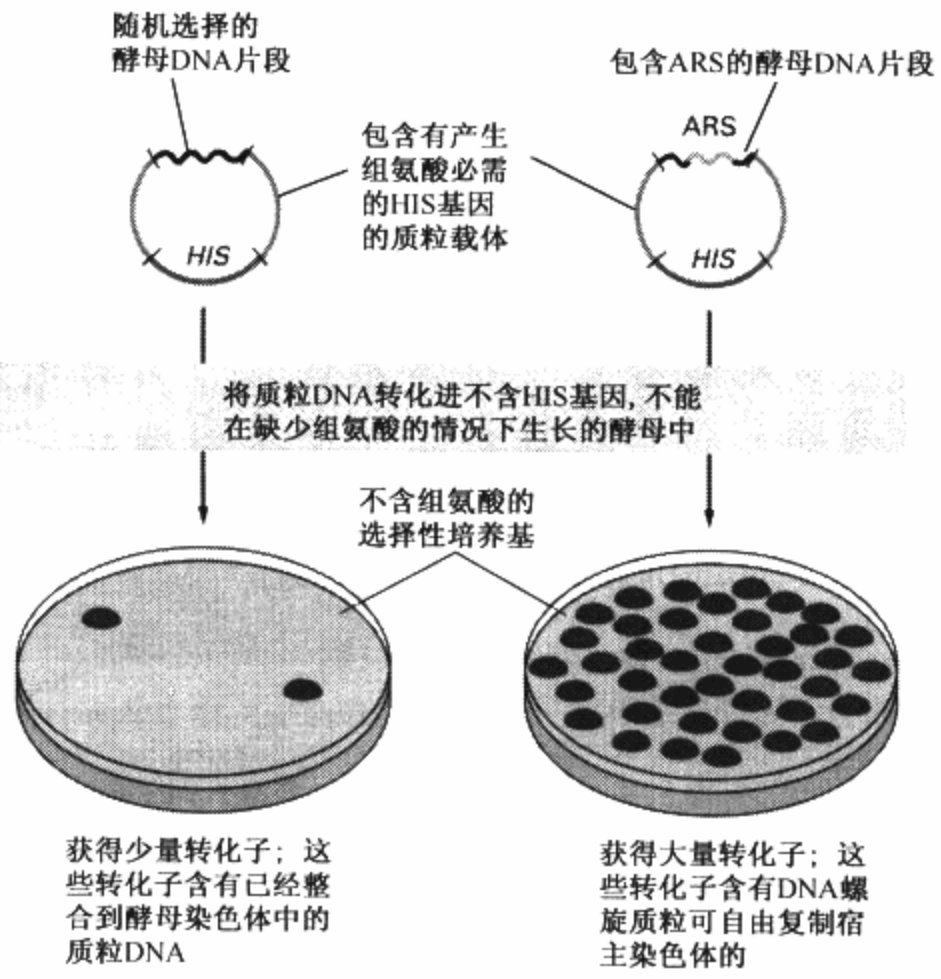


图 5-36　用于鉴定酵母细胞中复制起点的策略

以这种方式鉴定出的每一个酵母 DNA 序列称作自主复制序列（ARS），因为它使含有它的质粒能够在宿主细胞中复制，而不必整合到宿主染色体中

对于芽殖酵母，可以确定每个复制起点在每条染色体上的位置（图 5-37）。图 5-37 中显示的特定染色体（酿酒酵母 3 号染色体）还不到典型的人类染色体长度的 1/100。它的复制起点间隔 30 000 个核苷酸分布；起点的这个分布密度可以使酵母染色体复制只需约 8min。

在酿酒酵母中进行的遗传试验还检测了删除染色体Ⅲ上的各种不同复制起点后造成的影响。删除几个的复制起点影响较少，因为起始于相邻复制起点的复制能一直前进到缺少自身起点的区域内。然而，如果从该染色体删除更多的复制起点，这条染色体随机细胞分裂的进行就会慢慢丢失，可能是因为它复制得太慢。

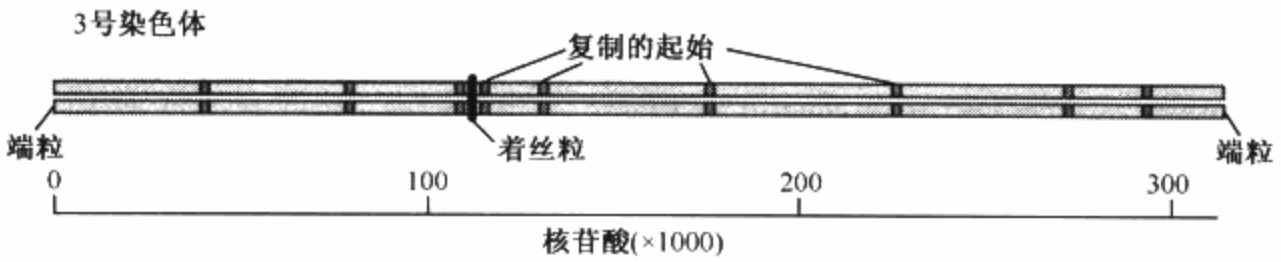


图 5-37　酿酒酵母 3 号染色体上的 DNA 复制起点

作为已知的最小真核生物染色体，它携带总共 180 个基因。如图所示，它含有 9 个复制起点

一个巨大的多亚基复合物结合在真核生物复制起点上

通过进行图 5-36 中显示的实验，利用越来越短的 DNA 片段，已经确定了酿酒酵母中指导 DNA 复制起始的最短 DNA 序列。酵母的每个复制起点都具有一个结合位点，结合一个巨大的多亚基起始蛋白，该起始蛋白称为 **ORC**，即**起点识别复合体**（origin recognition complex），也具有几个辅助的结合位点，结合帮助吸引 ORC 到 DNA 起点上的蛋白质（图 5-38）。

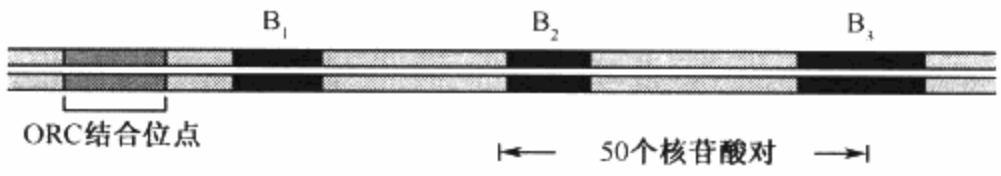


图 5-38　酵母的复制起点

含有大约 150 个核苷酸对的这个酵母复制起点（利用图 5-36 所示的程序鉴定得到）具有 1 个 ORC 结合位点，ORC 是一种结合每个复制起点的蛋白质复合物。图中所描述的起点也具有其他必需蛋白质的结合位点（B1、B2 和 B3），它所在不同的起点之间各不相同。虽然在酵母中已得到最好的鉴定，在更为复杂的真核生物中，包括人类，相似的 ORC 也用于起始 DNA 复制

我们已经看到，真核生物 DNA 复制只发生在 S 期。那么 DNA 复制是如何引发的呢？这种复制机制又是如何确保每轮细胞周期复制起点只使用一次呢？

这两个问题大致的答案现在我们已经清楚了，将在第 17 章讨论。简单地说，ORC 起点的相互作用是一种稳定的相互作用，在整个细胞周期内，可用作对复制起点进行标记。在 G_1 期，每个 ORC 上组装起一个前复制蛋白复合体，它含有一个六聚体的 DNA 解旋酶和一个解旋酶的装载因子（分别是 Mcm 和 Cdc6 蛋白）。当一种蛋白激酶活化时，就引发了 S 期，这个激酶将复制机器的剩余部分组装起来，使得一个 Mcm 解旋酶开始随复制起点处形成的任何一个复制叉移动。同时，这个引发 S 期的蛋白激酶阻止所有的 Mcm 蛋白进一步组装到前复制复合体中，直到这个激酶在下一个 M 期失活，以便重新开始下一轮周期（详细情况见图 17-22）。

哺乳动物中指定复制起始的 DNA 序列曾一度难以确定

与芽殖酵母中的情况相比，其他真核生物中指定复制起点的 DNA 序列难以定义。例如，对于人类，具有正确起点功能的 DNA 序列可以在 DNA 上延伸很大的距离。

然而，最近已经能够确定充当复制起点的特定人类 DNA 序列，每个有几千个核苷酸对长。运用重组 DNA 方法将这些起点转移到不同的染色区域后，它们继续发挥作用，只要是将它们置于染色体压缩程度较小的区域。这些起点序列中有一种来自 β-珠蛋白基因簇。当位

于基因组中正常的位置时，这个起点的功能关键地依赖于远距离的 DNA 序列（图 5-39）。在第 7 章中我们将会讨论。已经知道这个远距离的 DNA 序列对起点周围而且包括 β-珠蛋白基因在内的染色质结构具有解压缩的作用；很显然，结果产生的更加开放的染色质构型是这个起点正常发挥功能，以及 β-珠蛋白基因表达所必需的。

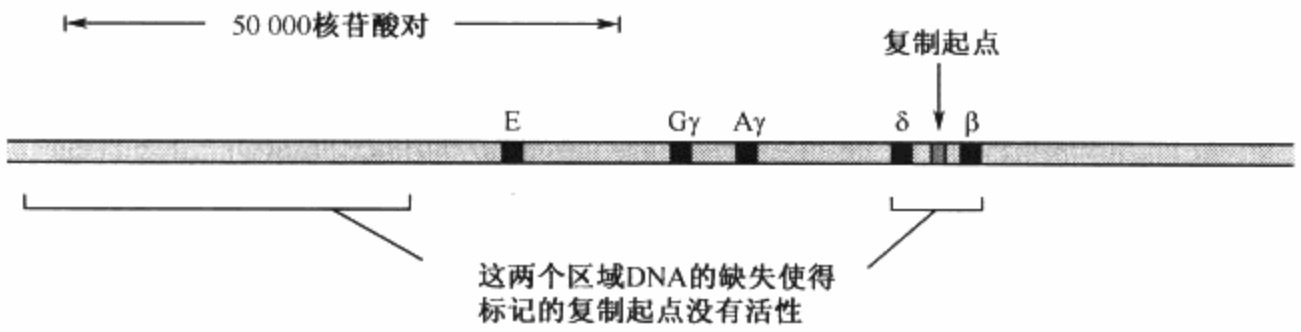


图 5-39　失活人类细胞中的复制起点的缺失

这两种缺失分别在两个患有地中海贫血病的个体中发现，这种病是由于不能表达图中所示的 β-珠蛋白基因簇中一个或多个基因而产生的紊乱。在这两种缺失突变体中，该区域的 DNA 由从 β-珠蛋白基因簇外的复制起点开始的复制叉进行复制，文中所解释的，左边的缺失去除了调控右边的复制起点染色质结构的 DNA 序列

现在我们知道，人类染色体复制起始需要一个与酵母细胞中相似的 ORC 复合体，而且同样的人类 Cdc6 和 Mcm 蛋白在起始过程中也起着关键作用。因此，可能酵母与人类复制起始机制看起来非常相似。然而，人类的 ORC 蛋白结合位点似乎没有哺乳酵母的特异，这可能解释了为什么人类复制起点比酵母的要更长，且定义更加宽泛。

新的核小体在复制叉后面组装起来

在这一部分以及下一部分，我们讨论真核生物特异的 DNA 复制过程中一些其他方面的问题。在第 4 章讲过，真核生物染色体由称作染色质的蛋白质与 DNA 的混合物构成。因此，染色体的复制不久需要拷贝 DNA，而且需要在每个复制叉后面的 DNA 上组装新的染色体蛋白质。尽管我们远未详细地弄清楚这个过程，但是我们正开始了解核小体是如何复制的，核小体是染色质包装的基本单位。在每个细胞周期，为了组装新的核小体，需要合成大量新的组蛋白，其质量差不多与新合成的 DNA 相等。因为如此，大部分真核生物的每种组蛋白的基因拥有大量的拷贝。例如，脊椎细胞有大约 29 个重复的基因群，大部分基因群含有编码所有的 5 种组蛋白的基因（*H1*、*H2A*、*H2B*、*H3* 和 *H4*）。

与大多数在整个间期持续合成的蛋白质不同，组蛋白主要在 S 期合成，这时由于转录量的增加而 mRNA 降解的量减少，组蛋白 mRNA 的水平增加约 50 倍。当 S 期末 DNA 合成停止时（或者当加入抑制剂过早地停止了 DNA 的合成时），通过依赖于 3′端的特殊性质的机制（讨论见第 7 章），主要的组蛋白 mRNA 变得极不稳定，并在几分钟内就会降解掉。相比之下，组蛋白自身却非常稳定，在细胞整个过程中都存在。DNA 合成和组蛋白合成的紧密偶联可能依赖于一种反馈机制，该反馈机制控制着游离组蛋白的水平，以确保合成的组蛋白数量与新合成的 DNA 的量恰好相当。

当复制叉前进时，它必须以某种方式穿过亲代核小体。体外研究表明复制装置具有一种我们了解甚少的内在能力，它能在通过亲代核小体时却不会改变它在 DNA 上的位置。第 4 章讨论的能使 DNA-组蛋白界面去稳定的染色质重建蛋白可能促进了细胞内的这个过程。

复制叉后的两个新合成的 DNA 双螺旋都继承了旧的组蛋白（图 5-40）。但是，由于 DNA 是增加了 1 倍，就需要增加相同量的新组蛋白，以完成 DNA 包装成染色质。向新合成的 DNA 上增加新的组蛋白受到染色质组装因子（CAF）的帮助，它是一类与复制叉结合的蛋白质，一旦在复制机器上出现就包装新合成 DNA。新合成的 H3 和 H4 组蛋白在它们的 N 端尾巴上迅速乙酰化（讨论见第 4 章）；当它们整合到染色质中以后，这些乙酰基团通过酶解作用从组蛋白上移走（图 5-41）。

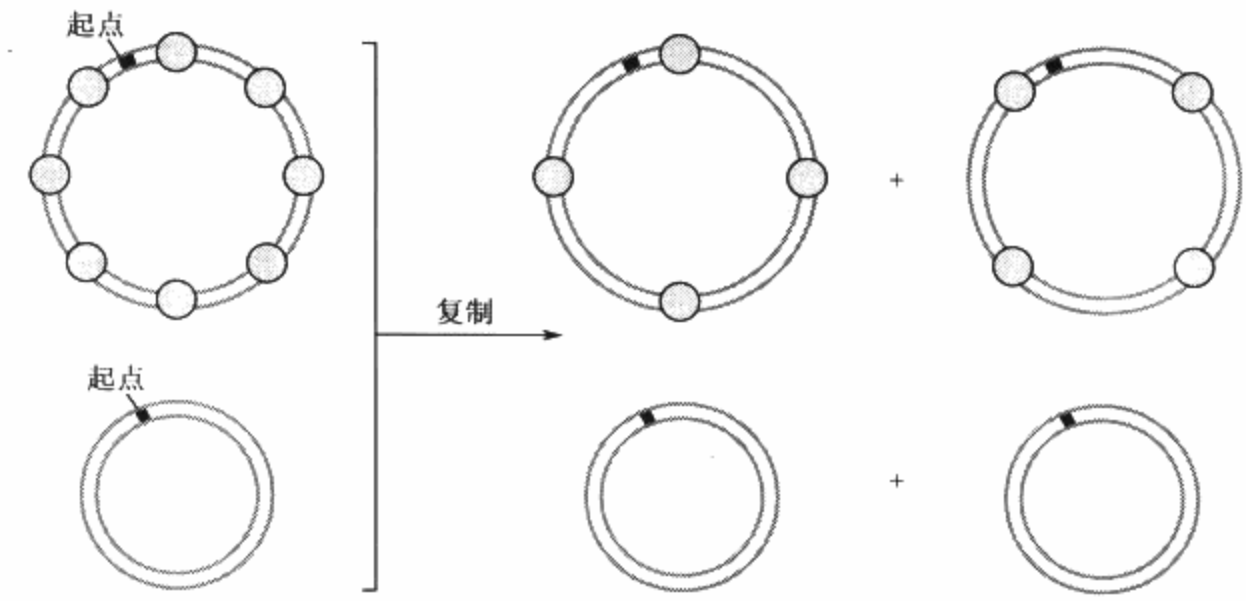


图 5-40　在复制叉通过后组蛋白保持与 DNA 结合的证明

在体外进行的该实验中，利用纯化的蛋白质复制两种大小不同的环状 DNA 分子混合物（其中只有一种装配成核小体）。经过一轮 DNA 复制，只有从形成核小体的亲本得到的子代 DNA 分子遗传了核小体。该实验也证明两个新合成的 DNA 螺旋都继承了原来的组蛋白

端粒酶复制染色体的末端

我们在前面看到，由于 DNA 聚合酶只能以 5′→3′方向聚合，所有在复制叉位置后随链的合成只能通过产生 DNA 短片段的倒缝机制，不连续地进行。当复制叉到达线性染色体末端时，该机制就遇到一个特殊问题：线性 DNA 分子的末端没有位置用来合成起始最后一个冈崎片段所需的 RNA 引物。

细菌由于染色体是环状 DNA 分子，所以不存在这个“末端复制”问题（图 5-30）。真核生物则以一种精确地方式解决了这个问题：它们的染色体末端有特殊的核酸序列，并整合到端粒内（讨论见第 4 章），吸引一种叫**端粒酶**（telomerase）的酶。端粒的 DNA 序列在各种不同的生物，如原生动物、真菌、植物和哺乳动物内都相似。它们由串联重复的短序列组成，这个序列中含有紧挨着排列的一组 G。在人类中，这个序列是 GGGTTA，并向两端延伸约 1000 个核苷酸。

端粒酶识别现有的端粒 DNA 重复序列的富含 G 的那条链末端，并将其向 5′→3′方向延伸。端粒酶利用自身组分中的 RNA 作为模板拷贝出新的重复序列。在其他方面下，端粒就像其他的逆转录酶，也就是利用 RNA 作为模板合成 DNA 的酶（图 5-42）。因此，端粒酶含有用于维持特有端粒序列的全部信息。通过端粒酶对亲代链的几轮延伸后，使用这些延伸的序列作为模板，在 DNA 聚合酶分子的作用下合成出互补链，从而完成染色体末端上的后随链的复制（图 5-43）。

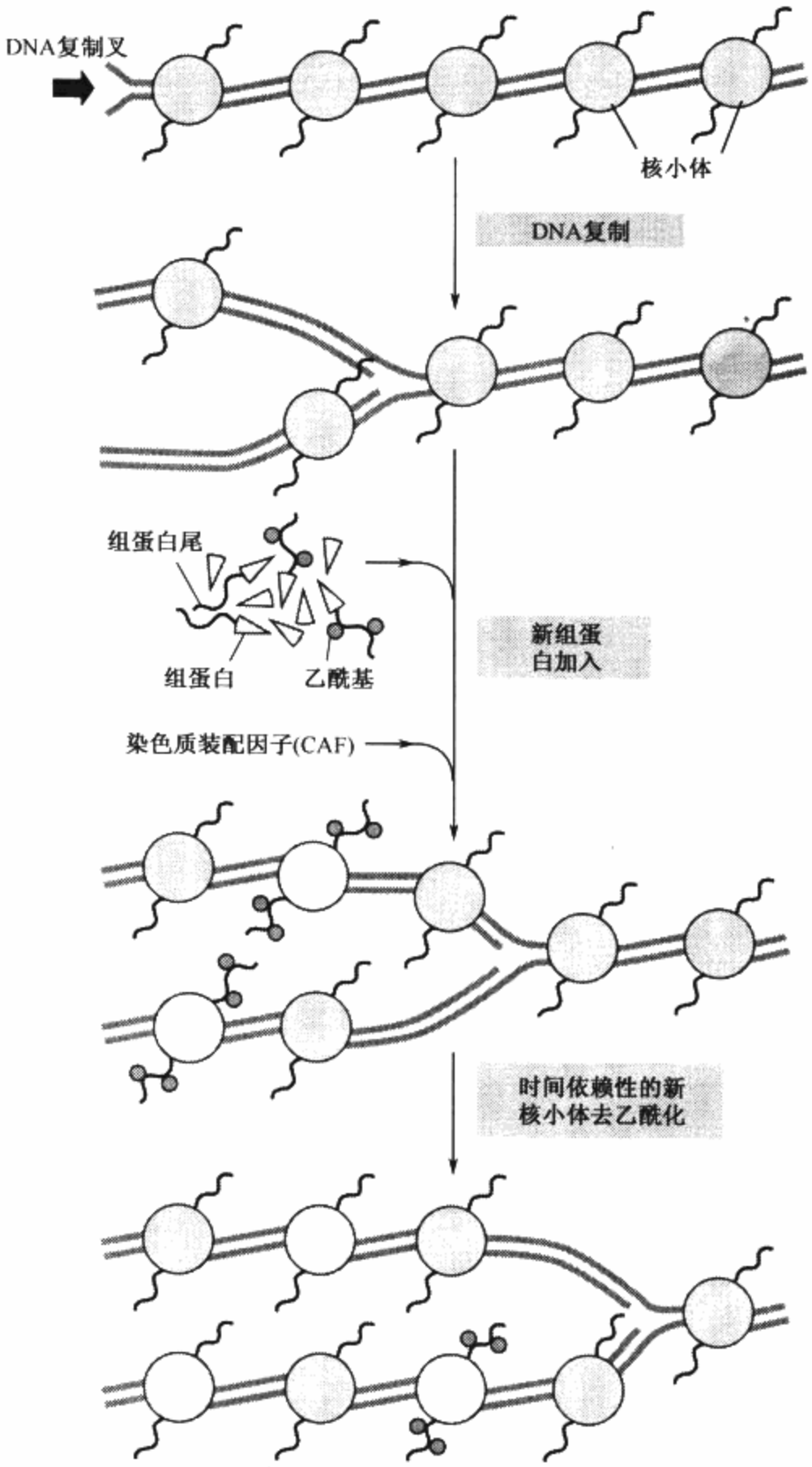


图 5-41　新组蛋白加入到复制叉后面的 DNA 上

该图中淡黄色的是新形成的核小体；如图所示，形成它们的组蛋白最初具有特异性乙酰化的赖氨酸侧链（图 4-35），该侧链后来被去除

上面描述的机制确保了每个端粒的 3′DNA 末端总是比与其配对的 5′端略长，形成一个突出的单链末端（图 5-43）。研究表明，在特殊蛋白质的帮助下，这个突出末端回折成环，将单链末端塞进端粒重复序列的双股 DNA 之间（图 5-44）。因此，染色体的正常末端具有一个独特的结构，该结构防止末端序列被降解，并且使它明显不同于细胞迅速修复的断裂 DNA 分子（图 5-53）。

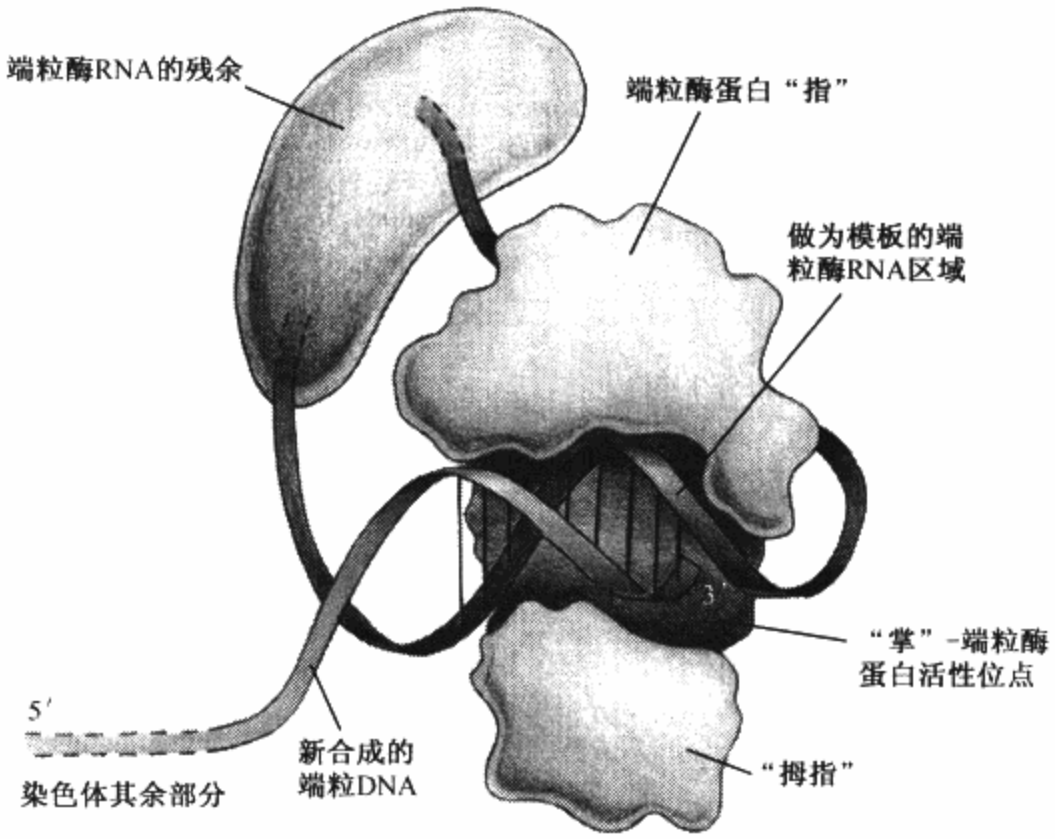


图 5-42　端粒酶的结构

端粒酶是一种携带有一条 RNA 模板用于合成一条重复的富含 G 的端粒 DNA 序列的蛋白质 RNA 复合物。这里只显示了与逆转录酶类似的端粒酶的一部分（绿色）。逆转录酶是一种特殊的利用 RNA 模板合成 DNA 链的聚合酶；端粒酶的独特之处在于它携带自身的 RNA 模板（摘自 J. Lingner and T. R. Cech, *Curr. Opin. Genet. Dev.* 8：226～232，1998）

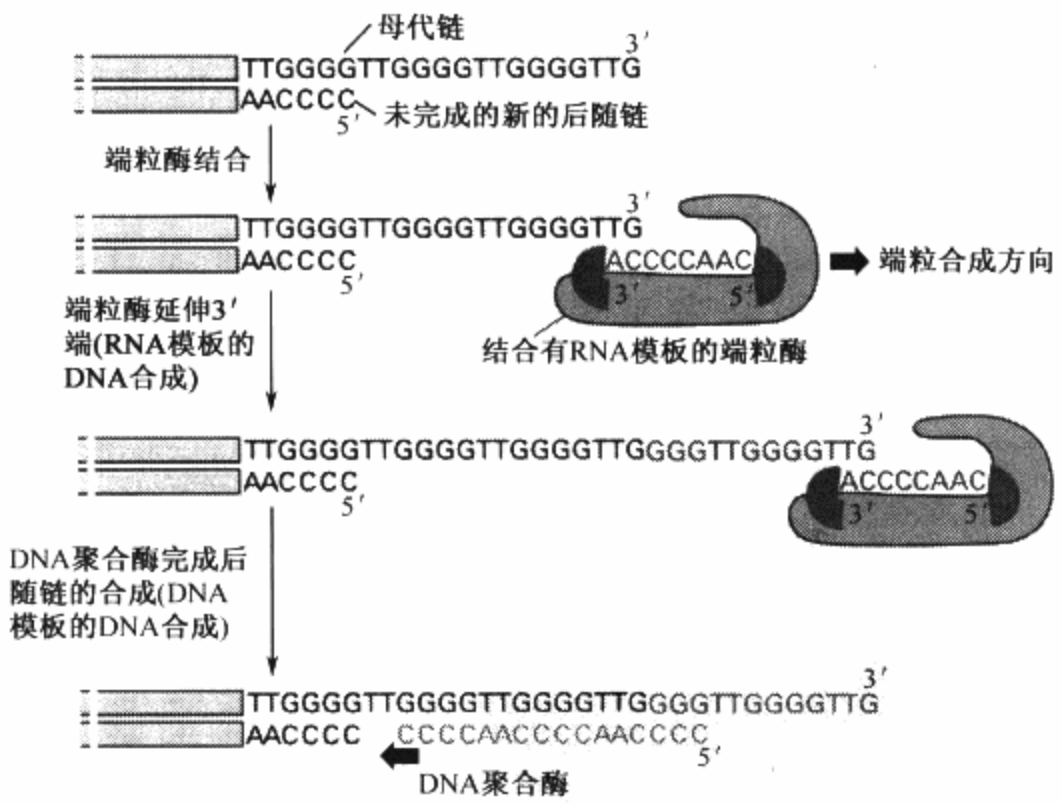


图 5-43　端粒的复制

这里显示的是参与合成重复的富含 G 序列的反应，这些序列形成各种真核生物染色体的末端（端粒）。亲本 DNA 链的 3′端由 RNA 为模板的 DNA 合成过程进行延伸；这使与其配对的不完整的子 DNA 链向 5′方向延伸。这条不完整的后随链可能由 DNA 聚合酶 α 完成，它携带一个 DNA 引发酶，作为其亚基之一（图 5-28）。图中所示端粒序列来自四膜虫，这些反应还是首次在这种动物中发现。在四膜虫中端粒重复是 GGGTTG，人类中是 GGGTTA，而酿酒酵母中是 GGGA

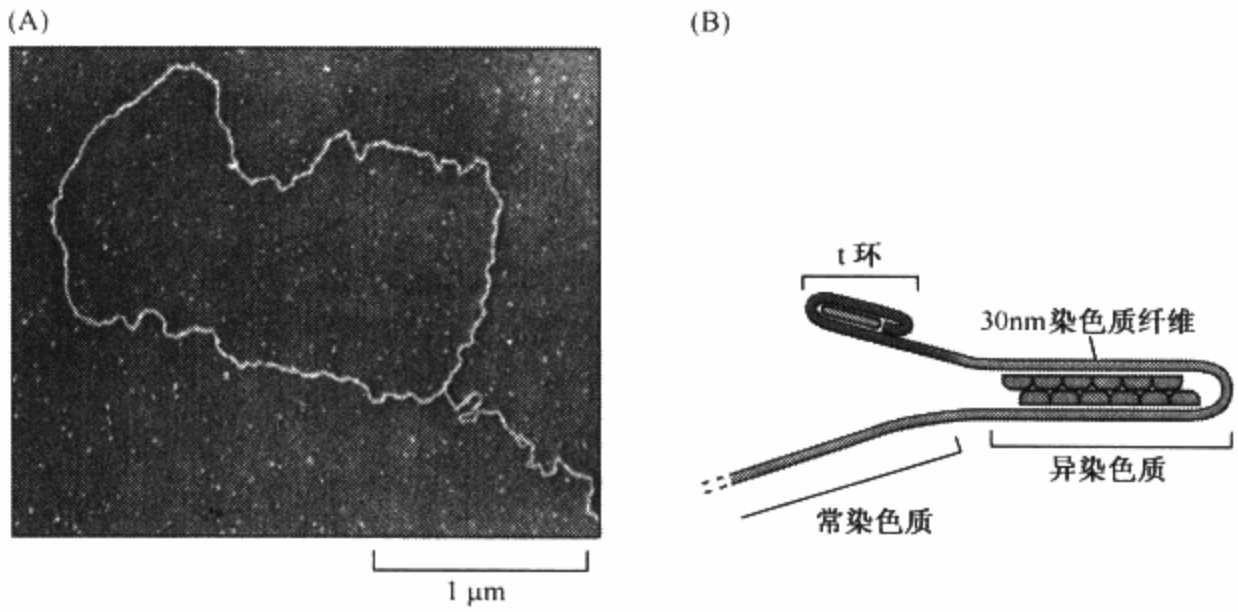


图 5-44　哺乳动物染色体末端的 t 环

（A）人的间期染色体末端 DNA 的电镜显微图。观察前将染色体固定，去除蛋白质，并人为加厚。这里所看到的环长度大约15 000个核苷酸对。（B）端粒结构的模型。特殊的蛋白质（用绿色标注）将单链末端插入到双链重复中以形成一个 t 环，并维持 t 环结构。此外，如图所示，有可能染色体末端通过 t 环附近的异染色质的形成再环绕一次(图4-47)[(A)摘自 J. D. Griffith et al., *cell* 97：503～514，1999. © Elseriver]

端粒长度受到细胞和生物体的调控

由于使每个端粒增长和缩短的过程只是近乎平衡，所以染色体的末端含有不同数目的端粒重复。毫不奇怪，实验表明无限增殖的细胞（如酵母细胞）具有自我稳定的机制，将这些重复的数目维持在一个有限的范围内（图 5-45）。

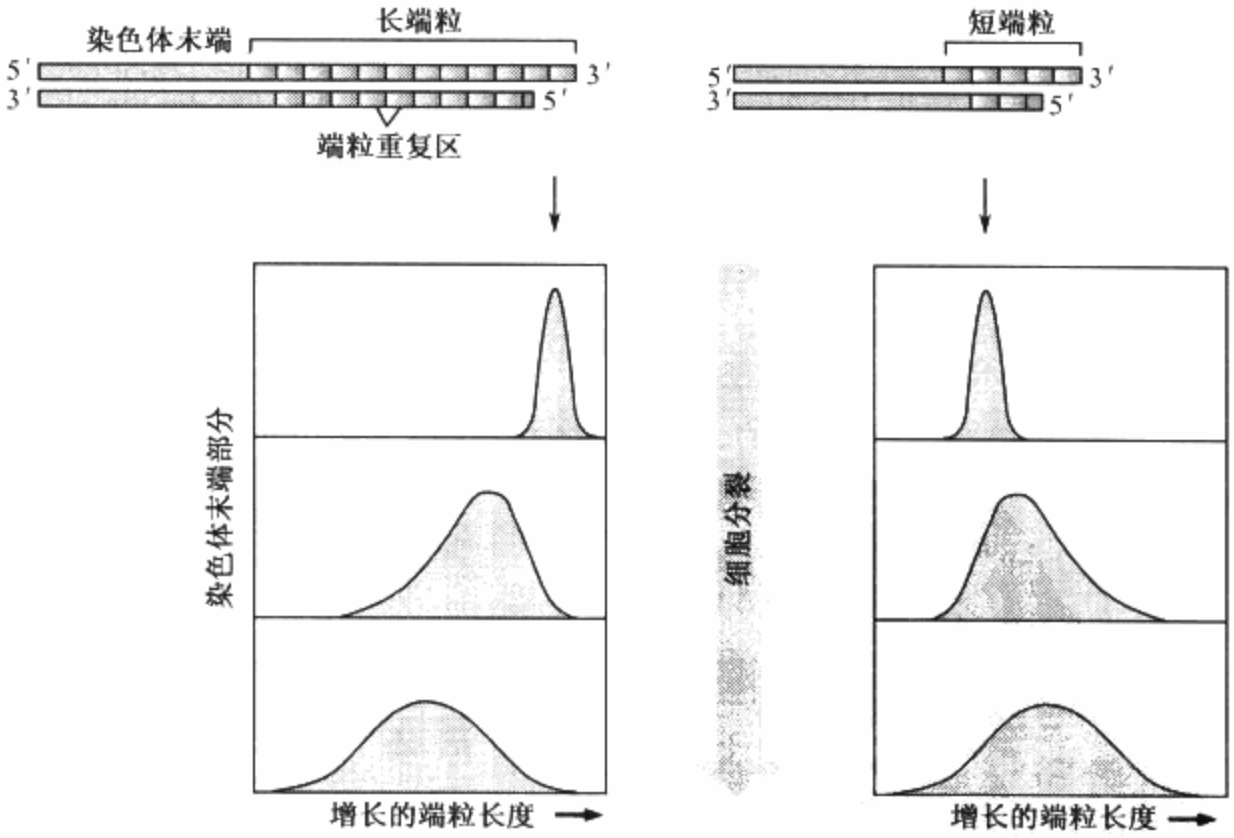


图 5-45　酵母细胞控制其端粒长度的证明

在该实验中，特定的染色体末端的端粒人为变得比平均大小或者长一些（左边）或者短一些（右边）。经过许多次的细胞分裂以后，染色体恢复原貌，显示出平均的端粒长度和长度分配是酵母细胞中其他染色体的典型特征。有人提出动物生殖系细胞中存在控制端粒长度的类似反馈机制

在人的体细胞中，端粒重复为每个细胞提供一种计数机制，帮助防止成年组织中异常细胞的无限增殖。根据这个观点，我们的体细胞一出生时具有完整的端粒重复；然而，在皮肤这样的组织中，端粒酶活性被关闭，所以每次细胞分裂时，就会从每个端粒丢失 50～100 个核苷酸。在许多代以后，后代细胞将会继承有缺陷的染色体（因为它们的末端不能完全复制），因此会永久性地停止细胞周期和分裂，这个过程称为复制细胞衰老（讨论见第 17 章）。理论上这样一种机制可以提供保护措施，防止人体组织中异常细胞不受控制的增殖，从而帮助我们防止癌症的发生。

端粒充当一根“测量棒”，用来计量细胞的分裂次数，从而调节细胞的生命，这个观点已从很多方面得到检验。对于某些在组织培养基中生长的人类细胞来讲，实验结果支持这种观点。培养基中的人成纤维细胞正常条件下经过 60 次细胞分裂增殖后，就会进入复制衰老。与人类其他大多数体细胞一样，成纤维细胞不能合成端粒酶，随着细胞的每次分裂，端粒就会逐渐变短。当通过插入一个活性的端粒酶基因，为成纤维细胞提供端粒酶时，端粒的长度就保持不变，并且许多细胞此时继续进行无限增殖。因此，看起来端粒变短显然能够在人类细胞中为细胞分裂计数，并引发复制衰老。

有人提出这种对细胞增殖的控制对于组织结构的维持很重要，而且它还以某种方式负责动物，如我们自己的衰老。通过培育缺少端粒酶的转基因鼠，这些观点得到了检验。小鼠染色体的端粒大约比人类的长 5 倍，因此在小鼠的端粒缩短到人端粒的正常长度之前必须将小鼠培育三代或三代以上。因此，毫不奇怪最初小鼠都正常地发育。更重要的是，后来的小鼠在它们高度增殖的组织中不断出现越来越多的缺陷。但是，这些小鼠看起来根本不会提早衰老，而且大一些的小鼠明显倾向于长肿瘤。在这些和其他的方面，这些小鼠与得遗传病先天性角化不良的人类似，这种病就是因为端粒过早缩短引起的。得这种病的人各种表皮结构（包括皮肤、趾甲和泪小管）和红细胞的制造都显示异常。

从上面的观察，可以清楚地知道，通过去除端粒控制细胞增殖为生物提供了一种冒险的尝试，因为在组织内不是所有缺少功能正常的端粒的细胞都会停止分裂。其他一些细胞在遗传上明显变得不稳定，但仍然持续分裂，就会产生各种导致癌症的细胞。因此，读者可以提出疑问，是否大多数人类体细胞缺少端粒酶的现象为我们提供了一种进化优势，正如推测端粒缩短倾向于保护我们不得癌症或其他增殖型疾病的人提出的那样。

小结

起始 DNA 复制的蛋白质结合到复制起点的 DNA 序列上，催化形成一个具有 2 个向外移动的复制叉的复制泡。当一个起始蛋白-DNA 复合体形成了，然后将 DNA 解旋酶放到 DNA 模板上的时候，这个过程就开始了，接着其他蛋白质结合上来，在每个复制叉上形成催化 DNA 合成的多酶“复制机器”。

在细菌和一些简单的真核生物中，复制起点规定为特异的 DNA 序列，只有几百个核苷酸对长。在其他的真核生物，例如，人类中，规定 DNA 复制起点所需的序列似乎并不确定，并且起点可以跨越几千个核苷酸对。

细菌的环状染色体上通常只有一个复制起点。由于复制叉的移动速度可达 1000 核苷酸/s，所以能在不到 1h 的时间内复制它们的基因组。真核生物 DNA 复制只在细胞周期的一段时期 S 期内进行。真核生物复制叉移动速度大约是原核生物的 1/10，而长得多的真核生物染色体每条都需要许多的复制起点，以便在通常持续 8h 的 S 期内完成复制。这些真核生物染色体上不同的复制起点按次序激活，激活的次序部分是由染色质结构决定，染色质结构压缩

程度高，则复制起始晚。复制叉经过以后，通过将新的组蛋白加入到直接以核小体的形式继承的旧的组蛋白中，重新形成染色质结构。

真核生物通过特化的末端结构——端粒，解决它们线性染色体末端复制的难题，这个过程需要一种特殊的酶端粒酶。端粒酶利用自身组成部分的 RNA 作为模板延伸端粒 DNA，产生高度重复的 DNA 序列，在染色体的每个末端通常可延伸长达 10 000 个核苷酸对或更长。

DNA 修复

尽管遗传的变化对进化很重要，但是个体的存活却要求遗传的稳定性。维持遗传的稳定不仅需要极其精确的 DNA 复制机制，而且需要修复 DNA 中不断出现的意外损伤的机制。DNA 中大部分的这种自发改变是暂时的，因为它们立即被共同称作 **DNA 修复**（DNA repair）的一系列过程加以校正。人类细胞 DNA 每天因热、代谢事故、各种辐射以及接触到环境中的多种物质而产生成千上万的随机改变中，只有少数作为突变在 DNA 中积累起来。现在我们知道 DNA 中 1000 个偶然碱基改变中，只有不到一个会导致永久性的突变，其余的都被 DNA 修复高度有效地清除掉。

DNA 修复的重要性从细胞质能合成如此多的 DNA 修复酶来看也很明显。例如，对细菌和酵母基因组进行分析后，发现这些生物体编码容量的百分之几都仅仅用于 DNA 修复功能。DNA 修复的重要性也可以通过 DNA 修复基因的突变后突变率的增加得到证明。许多 DNA 修复途径以及编码的基因（我们现在知道它们在各种各样的生物，包括人类中发挥作用）最初是通过在细菌中分离突变体并鉴定确定出来的，这些突变体表现出突变率或者对 DNA 损伤试剂敏感性的增加。

最近对人类 DNA 修复能力降低所造成的后果的研究将多种人类疾病与修复减少联系起来（表 5-2）。因此，如果一个基因在正常情况下主要功能是修复来源于 DNA 复制错误的错配碱基对，那么这个人类基因的缺陷就会导致可遗传的形成癌症的倾向，这反映了突变率的增加。在另一种人类疾病着色性干皮病，患者对紫外辐射极其敏感，因为他们不能修复某些 DNA 的光产物。这种修复缺陷导致突变升高，引起严重的皮肤伤害，以及更容易患某些癌症。

表 5-2　DNA 修复存在缺陷的遗传综合征

名称	表型	酶或过程缺陷
MSH2,3,6,MLH1,PMS2	结肠癌	错配
Xeroderma pigmentosum（XP）groups A—G	皮肤癌，细胞 UV 敏感，神经异常	碱基切除-修复
XP variant	细胞 UV 敏感	DNA 聚合酶翻译合成
Ataxia-telangiectasia(AT)	非白血性白血病、淋巴瘤，细胞 γ-射线敏感，基因组不稳定	ATM 蛋白，一种双链激活的蛋白酶损坏
BRCA-2	乳腺癌和子宫癌	同源重组修复
Werner syndrome	过早老化，多位点的癌症，基因组不稳定	附加的 3′-外切酶和 DNA 螺旋酶
Bloom syndrome	多位点癌症、发育障碍，基因组不稳定	为复制附加 DNA 螺旋酶
Fanconi anemia groups A—G	先天性异常，非白血性白血病，基因组不稳定	DNA 中间链交叉连接修复
46BR patient	对 DNA 损伤试剂超敏感，基因组不稳定	DNA 连接酶 I

如果不存在 DNA 修复，自发的 DNA 损伤将迅速改变 DNA 序列

虽然 DNA 是一种高度稳定的物质，正好符合储存遗传信息的要求，但是它是一种复杂的有机分子，即使在正常的细胞内条件下，也容易发生自发的改变，如果不加以修复，就会导致突变（图 5-46）。DNA 主要变化是热变动的结果。例如，每天每个人类细胞大约要丢失

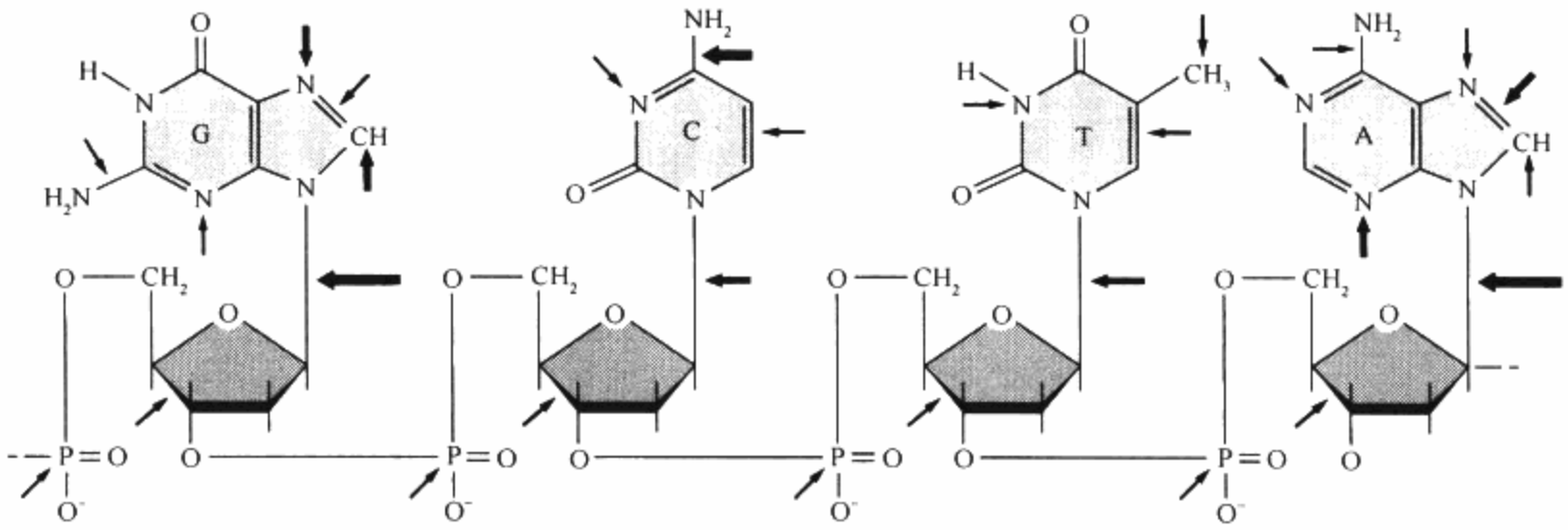


图 5-46　可能需要 DNA 修复的自发改变的总结

图中显示了每个核苷酸上可被修饰的位点，已知这些位点可能因自发的氧化损伤（红色箭头），水解作用（蓝色箭头）和甲基供体 S-腺甲硫氨酸（绿色箭头）不受控制的甲基化而修饰，每个箭头的宽度表示每种修饰发生的相对频率（引自 T. Lindahl. *Natune* 362：709～715，1993. © Macmillan Magazine Ltd）

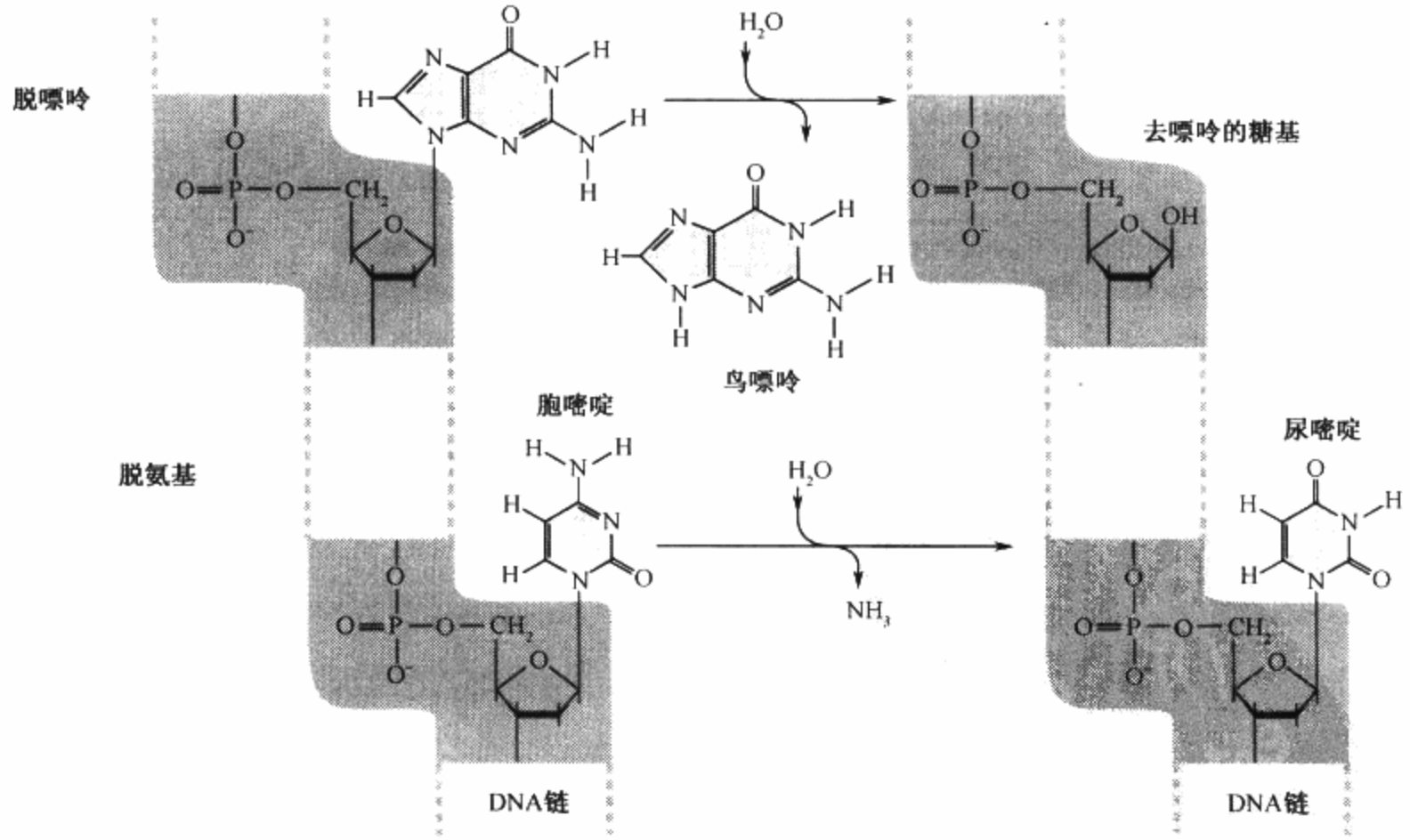


图 5-47　脱嘌呤和脱氨基

这两种反应是已知在细胞中引起严重 DNA 损伤的两种发生最频繁的自发的化学反应。脱嘌呤能从 DNA 释放鸟嘌呤（图中所示），还有腺嘌呤。主要类型的脱氨基反应（图中所示）将胞嘧啶转变成另一个改变的 DNA 碱基尿嘧啶，但脱氨基也在其他的碱基上发生。这些反应发生在双螺旋 DNA 上；为了方便起见，只显示了其中一条链

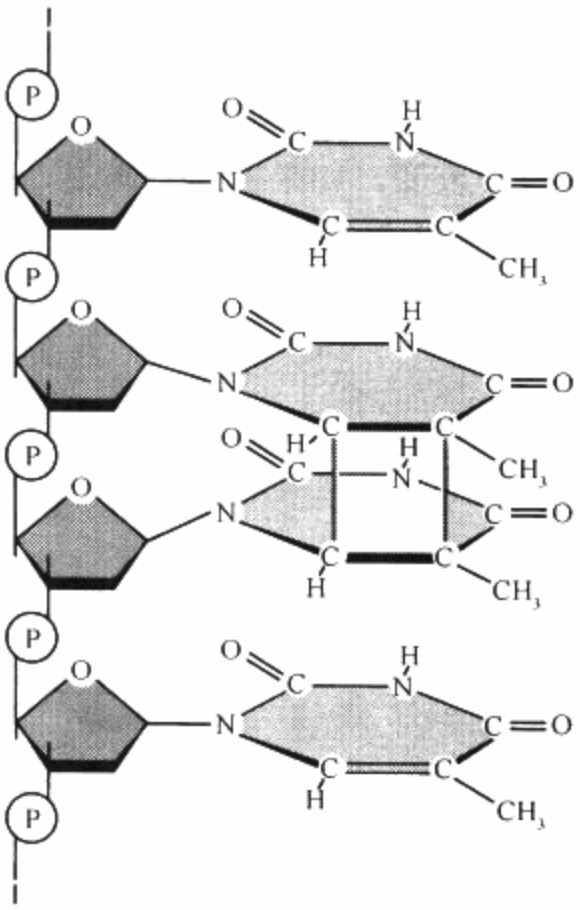


图 5-48　胸腺嘧啶二聚体

这种损伤会引入接触紫外辐射（比如阳光）的细胞内的 DNA。类似的二聚体会在 DNA 中任何两个相邻嘧啶碱基之间形成（C 或 T 残基）

5000 个嘌呤碱基（腺嘌呤和鸟嘌呤），因为它们与脱氧核糖之间的 *N*-糖苷键发生水解，这个反应称为脱嘌呤化。同样，每个细胞每天也以 100 碱基的速率自发在 DNA 中发生胞嘧啶的脱氨基化，变成尿嘧啶（图 5-47），DNA 碱基也因遇到活性代谢物（包括活性氧）或者环境中的化学物质而遭到破坏。同样，来自阳光中的紫外线也能在 DNA 中两个相邻嘧啶之间形成共价交联，如胸腺嘧啶二聚体（图 5-48）。在 DNA 复制时，如果没有加以校正，大部分的这些改变都会导致子代 DNA 链中缺失一个或多个碱基对，或者碱基对发生替换（图 5-49）。然后，这些突变在DNA 复制时就会通过后代细胞进行传播。这种高频率的随机改变将对生物体产生灾难性的后果。

DNA 双螺旋容易修复

DNA 双螺旋结构非常适合修复，因为它携带所有遗传信息的 2 个独立的拷贝——每条链上 1 个拷贝。因此，如果一条链损伤了，互补链仍然保留着相同信息的完整拷贝，并且这个拷贝常常用于恢复损伤链上正确的核苷酸序列。

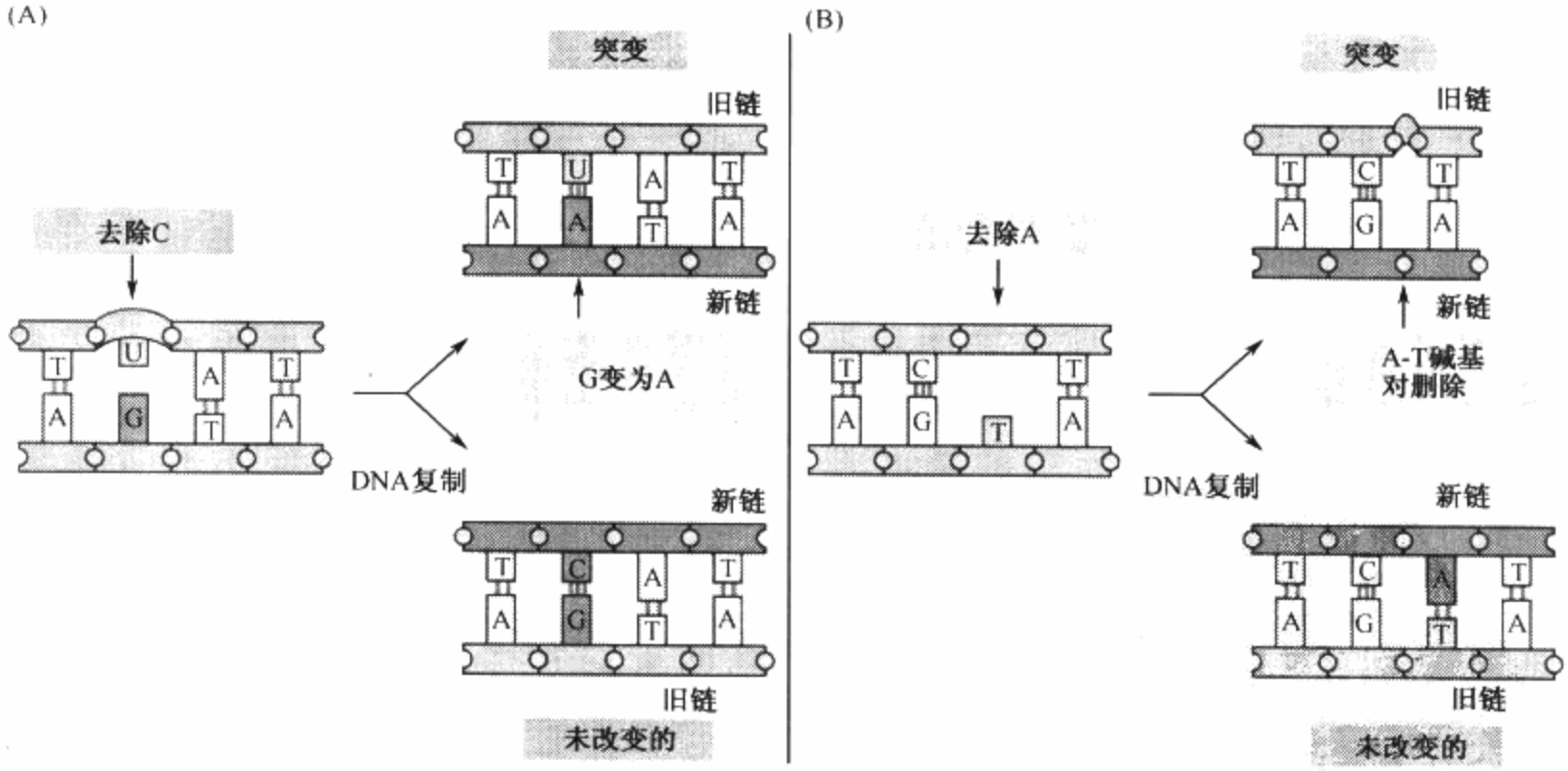


图 5-49　核苷酸的化学修饰如何形成突变

（A）胞嘧啶的脱氨基如果未加纠正，就会在 DNA 复制时引起一个碱基被另一个碱基替换。如图 5-47 所示，胞嘧啶脱氨基产生尿嘧啶。尿嘧啶与胞嘧啶的碱基配对特性不同，它优先与腺嘌呤配对。因此，DNA 复制机器遇到模板上的一个尿嘧啶时就加上一个腺嘌呤。（B）脱嘌呤如果未加校正，能引起一个核苷酸对的替换或丢失。当复制机器遇到模板链上缺失的嘌呤时，它可能跳到下一个完好的核苷酸上，如图所示，因此在新合成的链上产生一个核苷酸的缺失。许多其他类型的 DNA 损伤（图 5-46）如果没有校正，也会在 DNA 复制时产生突变

所有细胞都使用双股螺旋表明它对安全地储存遗传信息很重要，只有少数微小的病毒使用单链 DNA 或 RNA 作为它们的遗传物质。本部分描述的修复过程不能作用于这类核酸，因而在这些病毒单链基因组中发生永久核苷酸改变的概率非常高。看起来似乎只有具有小基因组的生物体才付得起代价，将它们的遗传信息编码在除 DNA 双螺旋以外的任何分子中。

每个细胞含有多种 DNA 修复系统，每个系统都有自己的酶以及识别损伤类型偏爱性。从本部分后面的内容我们可以看到，这些系统大都利用双螺旋未受损伤的链作为模板以修复损伤链。

DNA 损伤可以通过多种途径加以清除

存在多种 DNA 修复途径，它们使用不同的酶，作用于不同的损伤。图 5-50 显示了两种

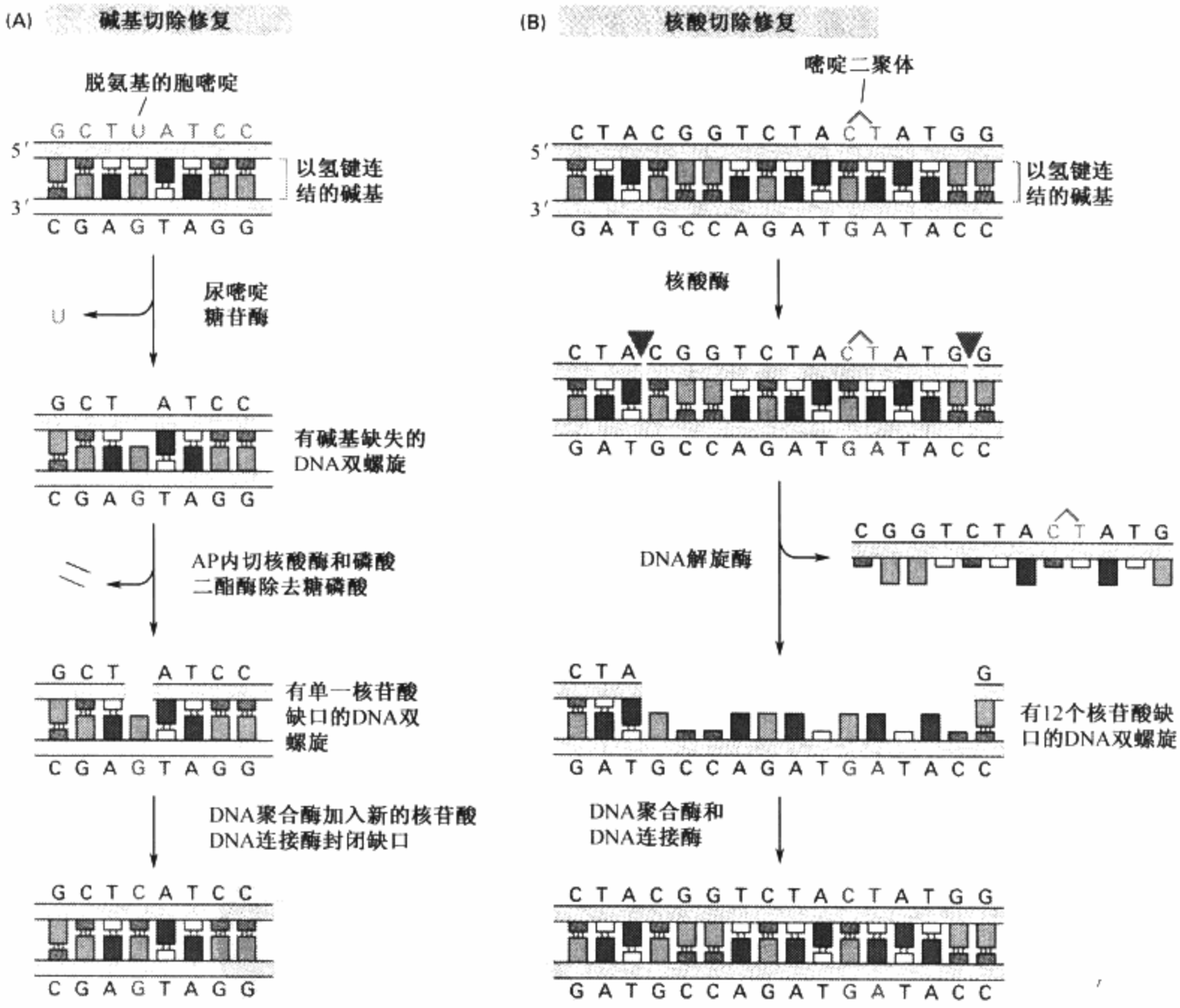


图 5-50　两种主要的 DNA 修复途径的比较

（A）碱基切除修复。该途径开始于一种 DNA 糖基化酶。这里尿嘧啶 DNA 糖苷酶去除 DNA 中的偶尔发生的脱氨基的胞嘧啶。在该糖苷酶（或识别一种不同的损伤的另一种 DNA 糖苷酶）发挥作用之后，ATP 内切核酸酶和一种磷酸二酯酶经过连续作用将缺少碱基的糖磷酸切除。（这些相同的酶直接修复脱嘌呤位点。）然后，DNA 聚合酶和 DNA 连接酶将此单核苷酸的缺口填平。结果就是将偶然脱氨基产生的 U 恢复成 C。AP 内切核酸酶名字的由来是因为它识别含有缺少一个碱基的脱氧核糖的 DNA 螺旋中任何位点；这种位点可因丢失嘌呤（脱嘌呤位点）或缺失嘧啶（脱嘧啶位点）而产生。（B）核苷酸切除修复。在一个多酶复合体识别诸如二聚体这样宽大的损伤后（图 5-48），就在损伤的两边切开，然后与之结合的 DNA 解旋酶将整个损伤链部分移开。细菌中的多酶复合体留下图中所示的 12 个核苷酸缺口；人类 DNA 中产生的缺口比这个要大 2 倍。核苷酸切除修复机制能识别并修复许多不同类型的 DNA 损伤

最常见的途径。在这两种途径中，损伤被切除，最初的 DNA 序列通过 DNA 聚合酶使用未受损伤的链作为模板得以恢复，双螺旋中剩余的断裂则由 DNA 连接酶进行封口（图 5-14）。

这两种途径的不同点在于从 DNA 移走损伤的方式不同。第一种途径，称为**碱基切除修复**（base excision repair），涉及一系列称为 DNA 糖苷酶的酶，它们每一种都能识别 DNA 发生变化的特定碱基，并催化将其水解去除。这些酶至少有 6 种，包括消除脱氨基的 Cs、脱氨基的 As、各种类型的碱化或氧化的碱基以及开环的碱基，碳碳双键意外地变成碳碳单键的碱基的酶。

图 5-50(A) 中显示的尿嘧啶 DNA 糖苷酶对脱氨基的胞嘧啶的去除，是碱基切除修复这种普遍机制的一个实例。发生改变的碱基是如何在双螺旋中被发现的呢？一个关键的步骤是由酶介导将改变的核苷酸从螺旋中“翻出来”，使得酶能够探测所有受损伤碱基的表面（图 5-51）。有人认为 DNA 糖苷酶使用碱基跳跃的方式沿 DNA 移动，对每个碱基对的状况进行评估。一旦发现一个损害碱基，DNA 糖苷酶就发动反应，生成一个缺少碱基的脱氧核糖。这个“缺掉牙齿”的部位被一种称为 AP 内切核酸酶的酶识别，该酶能切割磷酸二酯骨架，然后损伤被移走并修复[图5-50(A)]。脱嘌呤化是目前为止发现的 DNA 遭受的最频繁的一种损伤类型，它也能产生一个缺少碱基的脱氧核糖。脱嘌呤化经由 AP 内切核酸酶开始，以及紧接着的图 5-50（A）所示的后部分途径进行直接修复。

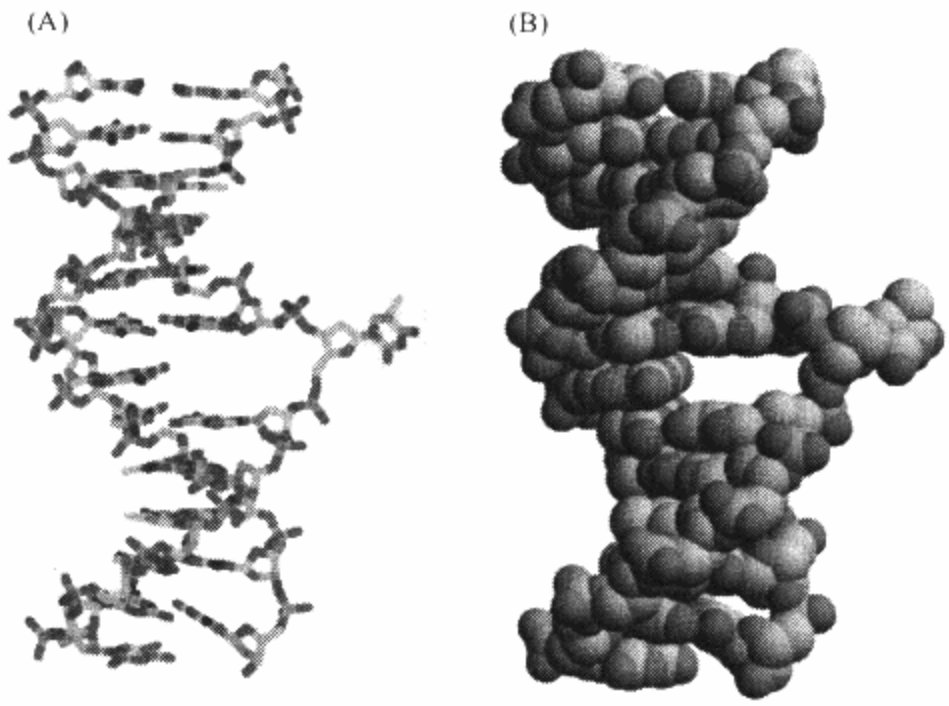


图 5-51　由碱基倒转所导致的 DNA 中异常碱基的识别

DNA 糖苷酶家族识别图中所示构象的特定碱基。这些酶中的每种酶切除连接特定识别的碱基（黄色）到糖骨架上的糖苷键，从而将其从 DNA 去除。（A）棒状模型；（B）空间充满模型

第二种主要的修复途径称为**核苷酸切除修复**（nucleotide excision repair）。这种机制可以修复 DNA 双螺旋结构内几乎任何的大改变所导致的损伤。这些“大损伤”包括大的碳氢化合物（如致癌物安息香比林）与 DNA 碱基发生共价反应的产物，以及阳光导致的各种嘧啶二聚体（T-T、T-C 和 C-C）。在这条路径中，存在一个巨大的多聚复合体扫描 DNA 双螺旋中的扭曲，而不是特定的碱基变化。一旦发现一个大损伤，异常链的磷酸二酯骨架上扭曲的两端就会切开，通过 DNA 解旋酶的作用含有这个损伤的寡核酸就会从 DNA 双螺旋脱离。然后 DNA 螺旋上形成的巨大的缺口就由 DNA 聚合酶和 DNA 连接酶进行修复[图5-50(B)]。

DNA 碱基的化学促进损伤的发现

DNA 双螺旋构建似乎最适合修复。上面提到过，由于 DNA 双螺旋含有遗传信息的备

份拷贝，所以如果一条链损伤了，另一条完好的链可以用作修复的模板。碱基的特性也促进区分受损和未受损的碱基。因此，每个可能发生的脱氨基化在 DNA 中都产生一个不正常的碱基，它能直接被识别，并被特异 DNA 糖苷酶清除。例如，次黄嘌呤是最简单的嘌呤碱基，它能够与 C 特异配对，但它是 A 的直接脱氨产物[图5-52(A)]。在次黄嘌呤上面再增加一个氨基就得到 G，它不通过自发脱氨基从 A 形成，而且它的脱氨基产物也同样是独特的。

正如第 6 章讨论的那样，从进化时间比例来考虑，人们认为 RNA 早在 DNA 之前，已经作为遗传物质。而且似乎遗传密码最初携带在 A、C、G 和 U 这 4 个核苷酸中。这引发了另一个问题，为什么 RNA 中的 U 在 DNA 中被 T（T 是与 5-甲基 U）取代了呢？我们已经看到 C 的自发脱氨基将其转化为 U，但是这个事件因尿嘧啶 DNA 糖苷酶作用变得相对无害。然而，如果 DNA 将 U 作为正常的碱基，修复系统将难以区分脱氨基的 C 和自然存在的 U。

在脊椎动物 DNA 中存在一种特殊情况，特殊的 C-G 序列中的 C 核苷酸被甲基化，这些被甲基化的序列与失活的基因相关（讨论见第 7 章）。这些甲基化的 C 偶尔发生脱氨基化，生成自然的核苷酸 T[图5-52(B)]，与互补 DNA 链上的 G 形成错配。为了帮助修复脱氨基的 C 核苷酸，一种特殊的 DNA 糖苷酶识别序列 T-G 中的涉及 T 的错配碱基对，并去除 T。然而，这种修复机制必须比较有效，因为甲基化的 C 是脊椎动 DNA 常常发生突变的位点。引人注意的是，尽管人类 DNA 中的 C 核苷酸仅有 3%的发生甲基化，但是这些甲基化的核苷酸中发生的突变约占人类遗传疾病中已观察到的单碱基突变的 1/3。

双链断裂被有效地修复

一种具有潜在危险的 DNA 损伤类型是双螺旋两条链都发生断裂，使得没有完好的模链用于修复。这种类型的断裂是由离子辐射、氧化试剂、复制错误，或者细胞内某些代谢产物引起的。如果这些损伤未加以修复，它们将迅速导致染色体断裂成更小的片段。然而，细胞进化出两种不同的机制用于修复潜在的损伤。最容易理解的一种是非同源末端连接，在这种机制中把断裂的末端并列后由 DNA 连接酶重新连接起来，同时往往在连接位点丢失一个或多个核苷酸[图5-53(A)]。这种末端连接机制，可以看作是修复双链断裂的一种紧急措施，是哺乳动物细胞内常见结果。尽管在断裂位点导致 DNA 序列发生改变（一个突变），但是哺乳动物基因组只有很小一部分编码蛋白质，以至于这种机制显然是一种可接受的措施，用于保证染色体完整性。前面讨论过，端粒末端的特殊结构防止染色体末端被误认为是断裂的 DNA，因此保护了自然的 DNA 末端。

一种甚至更加有效的双链断裂修复利用了一个事实，就是双倍体细胞含有每个双螺旋的两个拷贝。在这第二种称作同源末端连接的修复途径中，一般性重组在其中发挥作用，它将核苷酸序列信息从完整的 DNA 双螺旋上转移到断裂螺旋中的双链断裂处。这种反应需要特殊的重组蛋白，它能识别两条染色体之间配对的 DNA 序列，并使其靠近。然后，DNA 修复过程利用完好链作为模板，转移遗传信息到断裂的染色体，将其修复且不改变其 DNA 序列[图5-53(B)]。在那些 DNA 已修复但还未分裂的细胞中，在每条染色体的两条姐妹 DNA 分子之间能迅速地进行这种修复；在这种情况下，就不需要断裂的末端在同源染色体中寻找配对的 DNA 序列。对于这种同源末端连接反应更加详细的分子机制将在本章后面讨论，因为它要求对细胞进行遗传信息重组的方式有一个大致的了解。除在人类中存在以外，这种 DNA 双链断裂修复方式也是细菌、酵母和果蝇，所有几乎未观察到非同源 DNA 末端的生物中最重要的修复方式。

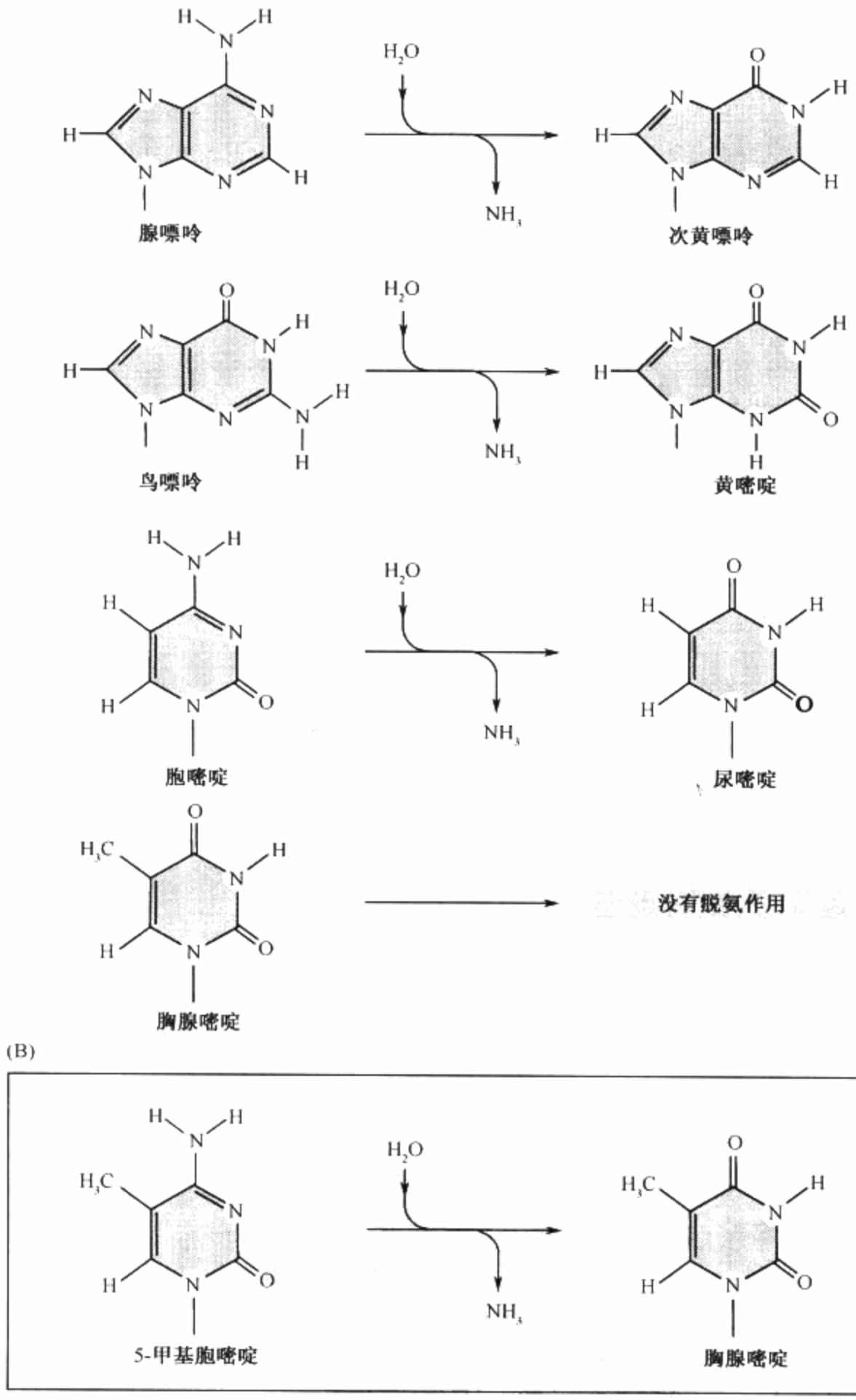


图 5-52　DNA 核苷酸的脱氨基

在每个例子中该反应中随水加入的氧原子用红色表示。(A) 当 DNA 发生 A 和 G 自发脱氨基化时，其产物作为异常物加以识别，因此容易被识别和修复。C 和 U 的脱氨基已在前面图 5-47 说明；T 没有氨基，不能用于脱氨基化。(B) 脊椎动物 DNA 中大约 3%的 C 核苷酸都甲基化，以帮助调控基因表达（讨论见第 7 章）。当这些 5-甲基胞嘧啶核苷酸偶尔发生脱氨基时，它形成自然的核苷酸 T。这个 T 将与相反链上的 G 配对，形成一个错配的碱基对

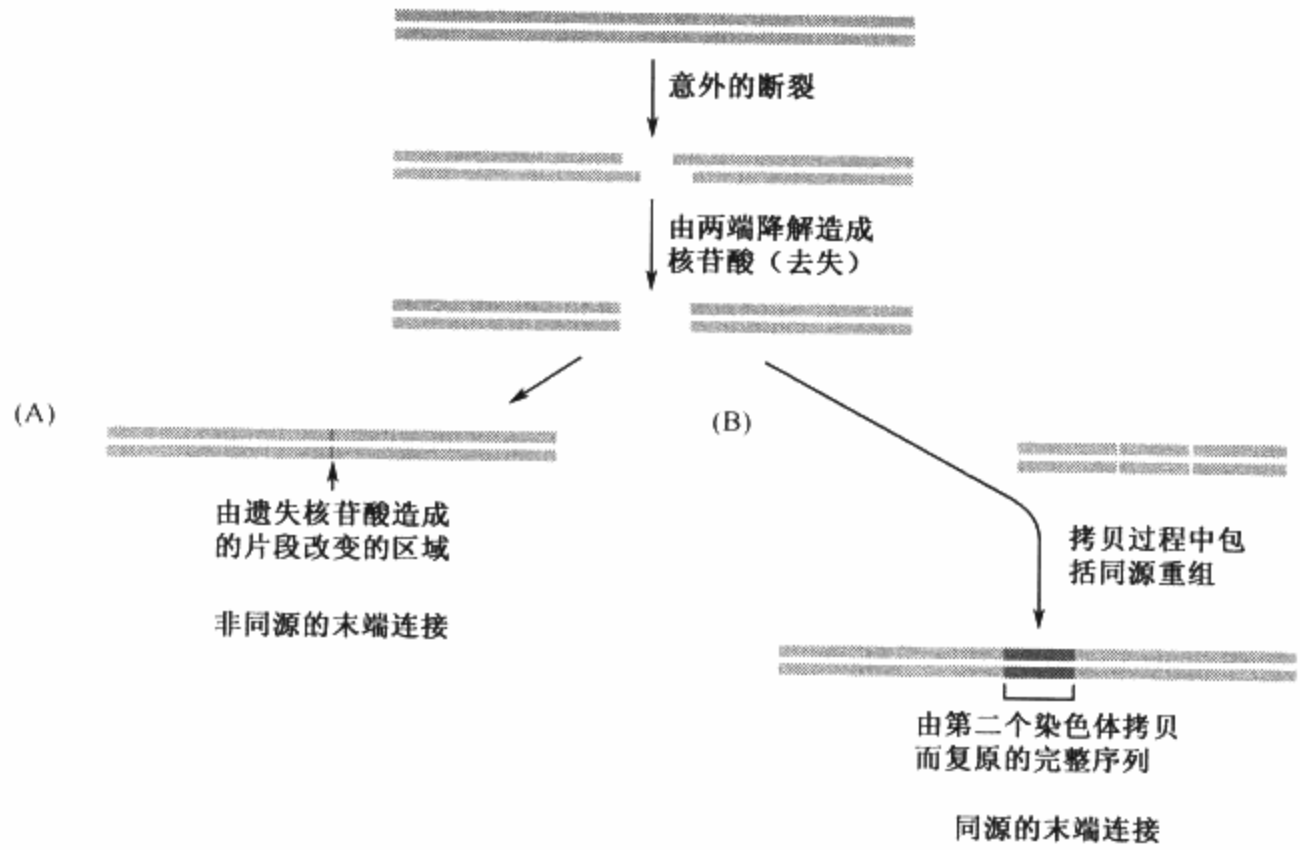


图 5-53　修复双链断裂的两种不同类型的末端连接

（A）当修复打断染色体时，非同源的末端连接改变原先 DNA 序列。这些改变可以是缺失（如图所示）或者短片段插入。（B）同源的末端连接更难完成，但要精确得多

在对 DNA 损伤的应答中细胞能产生 DNA 修复酶

细胞已经进化出许多机制，帮助它所在难以预料的危险环境中存活。细胞环境的极端变化经常会激活一系列基因的表达，它们的蛋白质产物能保护细胞不受这种变化的有害影响。所有细胞都共同存在的一种这样的机制就是热激反应，它因细胞暴露到异常高温时激发。诱导产生的“热激蛋白”包括某些帮助稳定和修复部分变性细胞的蛋白质，这些在第 6 章讨论。

细胞还拥有一种机制，能增加 DNA 修复酶的水平，以作为对严重 DNA 损伤的紧急反应。研究最清楚的例子就是大肠杆菌中的所谓的 **SOS 反应**（SOS response）。在这种细菌内，任何因 DNA 损伤引起的复制的阻断，都会产生信号，诱导至少 15 种基因的转录增加，它们中许多都编码在 DNA 修复中发挥作用的蛋白质。这种信号（认为是过量的单链 DNA）激活大肠杆菌 RecA 蛋白质（图 5-58），这样它就会破坏一种基因调控蛋白，该蛋白质正常情况下抑制大量 SOS 反应基因的表达。

对 SOS 反应过程中不同部分缺陷的细菌突变进行研究后证明，新合成的蛋白质有两种效应。首先，正如我们推测的，这些另外的 DNA 修复酶的诱导增加了细胞在 DNA 损伤后的存活率。其次，诱导蛋白中有些通过增加 DNA 复制过程的错误量短暂地增加了突变率。这些错误的产生是因为合成了低忠实度的 DNA 聚合酶，它利用损伤的 DNA 作为 DNA 合成模板。尽管这种“错误倾向”的 DNA 修复可能对单个细菌有害，但是估计这从长远看也具有好处，因为它在细菌群落中产生突发的遗传可变性，增加了产生更能在改变的环境中生存的突变体细胞的可能性。

人类细胞至少含有 10 种小的 DNA 聚合酶，它们中许多作为最后的手段特异性地参与复制 DNA 模板中未修复的损伤处。这些聚合酶分子每一种都有机会加入一个或多个核苷酸，因为这些酶在拷贝正常 DNA 序列时都是极其倾向出错的。尽管有关这些令人着迷的反应的

细节还未完全弄清楚，但是它们提供了完美的证据，表明生物体在精心地维护它们的 DNA 序列。

DNA 损伤延迟了细胞周期的进程

我们刚才看到了细胞含有多种酶系统，它们能识别 DNA 损伤并促进这些损伤的修复。由于代代相传时维持完好未受损的 DNA 的重要性，细胞具有了另一种机制帮助它们应对 DNA 损伤：它们能延迟细胞周期的进程直到 DNA 完成了修复。例如，在应答大肠杆菌 SOS 信号时表达的基因有一种是 sulA，它编码一种细胞分裂的抑制物。因此，当在对 DNA 损伤做的应答中细胞开启 SOS 功能时，就会阻碍细胞分裂，以延长时间供修复。当 DNA 修复完成时，SOS 基因的表达受到抑制，细胞周期继续进行，完好的 DNA 就会分离到子代细胞中。

真核生物中损伤的 DNA 也产生信号阻断细胞周期的进程。第 17 章将详细地讨论，细胞周期有条不紊地进行是通过利用“检查点”，确保在开始下一步之前，上一步已完成了。在这些细胞周期检查点中的几个检查点，如果发现损伤 DNA 细胞周期会停下来。因此，在酵母中 DNA 损伤的存在会阻止细胞周期进入 G_1 期；DNA 复制一旦开始就会减慢速度；而且它还能阻止 S 期向 M 期的过渡。DNA 损伤导致许多 DNA 修复酶的合成增加，而周期的延迟为修复的完成提供所需的时间，因而进一步促进了修复。

这些对 DNA 损伤做出应答的特殊信号转导机制的重要性，从编码 ATM 蛋白（一种大的蛋白激酶）基因存在缺陷的出生的人的表型中得到说明。这些个体患有共济失调毛细管扩张（AT），其症状包括神经退行性变，容易得癌症，以及基因组不稳定。在人类和酵母中，都需要 ATM 蛋白质来生成最初的细胞内信号，对氧化导致的 DNA 损伤做出应答。这种蛋白质存在缺陷的生物个体对于引起这类损伤的试剂，如离子辐射，表现出超敏感性。

小结

遗传信息能够在 DNA 序列中稳定储存，只是因为大量的 DNA 修复酶不断地扫描 DNA 并置换出任何受损的核苷酸。大多数类型的 DNA 修复依赖于 DNA 双螺旋每条链上都存在遗传信息单独的一份拷贝。因此，一条链上偶尔的损伤能够被修复酶切除，并通过参考完好链中的信息合成出正确的链。

大部分的 DNA 碱基损伤通过两条主要的 DNA 修复途径之一加以切除。在碱基切除修复中，改变的碱基由 DNA 糖苷酶移走，再将所产生的糖磷酸进行切割。在核苷酸切除修复中，损伤周围 DNA 链的一小部分从 DNA 双螺旋中作为寡核苷酸移走。在两种情况下，在 DNA 上留下的缺口通过 DNA 聚合酶和 DNA 连接酶利用未受损的 DNA 链作为模板，连续作用后填补完整。

其他一些关键的修复系统，基于非同源或同源末端连接机制，将 DNA 螺旋中偶然发生的双链重新封闭起来。在大多数细胞中，DNA 损伤水平的增高不仅引起修复酶增加，还会造成细胞周期的延迟。这两种因素帮助保证在细胞分裂之前，DNA 损伤得以修复。

一般性重组

在前面两个部分，我们讨论了细胞内维持 DNA 序列代代相传时基本不变的机制。然而另一方面也很明显，就是这些 DNA 序列偶尔能重新排列。任何个体基因组中基因的特定组合，以及这些基因表达的时间和水平，都常常因这种 DNA 重排发生改变。在一个群体内，

这类遗传变化对于使生物能够相应于改变的环境进行进化很关键。DNA 重排是由一系列机制引起的，这些机制一起称为**遗传重组**（genetic recombination）。通常认为存在两类广泛意义上的重组——一般性重组和位点特异性重组。本章这一部分我们首先讨论第一种机制，下一部分我们讨论第二种。

在**一般性重组**（general recombination）中（也称为同源重组），遗传交换发生在一对同源 DNA 序列之间。它们通常位于相同染色体的两份拷贝上，尽管具有相同核苷酸序列的其他 DNA 分子也能够参与。一般性重组反应对于每个增殖细胞都很重要，因为几乎每一轮 DNA 复制都可能发生事故，干扰复制叉前进，因而需要一般性重组机制进行修复。有关复制与重组之间的紧密相互作用的细节仍然没有完全了解，但是其中包括了使用同源末端连接反应（图 5-53）的变种，以重新起动在亲代 DNA 模板遇上断裂而停止的复制叉。

一般性重组对于真菌、植物和动物中有丝分裂过程中染色体的准确分离也很重要（图 20-11）。其结果产生染色体的交联，进一步导致部分遗传信息发生交换，并在每条染色体上形成 DNA 序列的重新组合。这种基因混合对于进化的好处是如此的显而易见，以至于由一般性重组引起的基因重排不只局限于多细胞生物内发生，在单细胞生物中也广泛存在。

一般性重组机制的本质特征似乎在所有生物体中都相似。我们所了解的有关遗传重组化学知识中，大部分最初都来源于对细菌，特别是大肠杆菌及其病毒的研究，还有在简单真核生物中如酵母中进行的实验。对于这些世代时间短，基因组相对较小的生物来讲，有可能分离出大量在重组过程存在缺陷的突变体。然后对每种突变体中发生改变的蛋白质进行鉴定后，就能鉴定和描绘特征催化一般性重组的蛋白质的集合。最近，已经在果蝇、小鼠和人类中发现了这些蛋白质的同源物，并对它们的特征做了大量的描述。

一般性重组由两个同源 DNA 分子间碱基配对互作指导进行

有丝分裂中观察到的广泛的一般性重组具有以下特征：①原本位于不同染色体上的两个同源 DNA 分子发生“交换”，也就是说，它们的双螺旋发生断裂，然后两个断裂的末端与相反的同源链连接起来，重新形成两条完好的双螺旋，每个都由两条最初的 DNA 分子组成（图 5-54）。②交换的位点（也就是图 5-54 中红色双螺旋与绿色双螺旋连接的位点）可能位于两条参与重组的 DNA 分子的同源核酸序列中任何位置。③在发生交换的位点，其中一个 DNA 分子上的一条链与另一 DNA 分子上的一条链形成碱基配对，产生一个异源双链连接，将两个双螺旋连接起来（图 5-55）。这个异源双链区域可以长达几千个碱基对；我们将在后面解释它是如何形成的。④在交换位点没有核苷酸序列发生改变；通常也进行一些 DNA 复制，但是断裂和再连的过程如此精确，以至于没有任何一个核苷酸丢失或增加。除了精确以外，一般性重组形成具有新序列的 DNA 分子：异源双链连接能够容纳一小段错配的碱基对，更重要的是，这两个交叉的 DNA 分子，在发生连接的两个末端常常不一样。因此，就形成了重组后新的 DNA 分子（重组染色体）。

一般性重组机制确保了两个 DNA 双螺旋只有当它们含有一段序列相似（同源）的区域时，才能够进行交换反应。一条长的异源双链连接的形成要求这种同源性，因为它涉及最初两个双螺旋互补的两条链形成的一长段互补配对区域。但是，这个异源双链连接是怎样产生的以及交叉位点上两个同源区域的 DNA 是怎样彼此识别的呢？我们将会看到，识别发生在称为 **DNA 联会**（DNA synapsis）的过程。在此期间来自两条 DNA 分子的互补链之间形成碱基配对。接着这个碱基配对进行延伸，指导一般性重组过程的进行，使其在含有长段配对（或几乎配对的）DNA 序列区域的两个 DNA 分子之间发生。

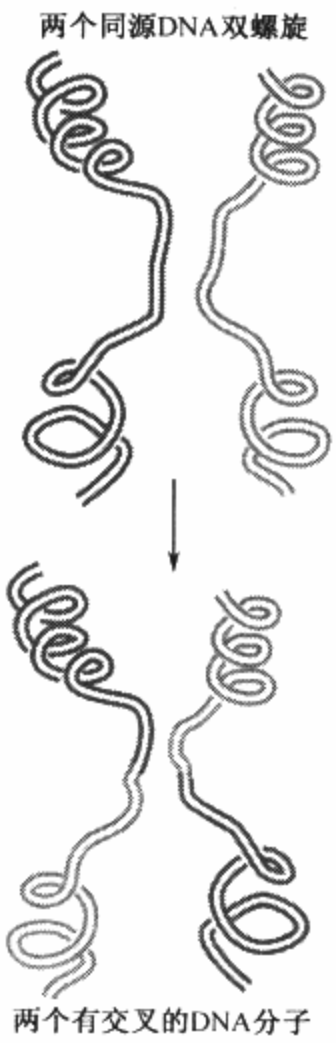


图 5-54　一般性重组

两条 DNA 双螺旋的断裂和再连接产生两个“交叉”的 DNA 分子。在减数分裂中，这个过程引起生殖细胞中每条染色体含有从父亲和母亲遗传的基因混合物

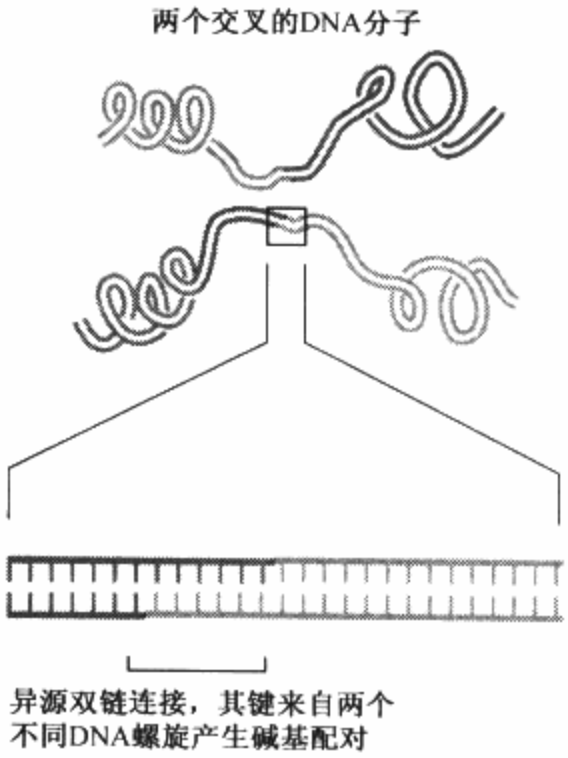


图 5-55　异源双链连接

该结构将两个发生交换的 DNA 分子连接起来。这种连接通常长达数千核苷酸

减数分裂重组由双链 DNA 断裂引起

两条完好的 DNA 双螺旋之间并不能发生大量的碱基配对相互作用。因此，对减数分裂期间发生一般性重组很关键的 DNA 联会只有在来自其中一个 DNA 螺旋的一条链已经暴露，且它的核苷酸能够与另一个 DNA 螺旋发生配对之后才能开始进行。在缺乏直接的实验证据时，人们提出许多理论模型，它们基于这样的观点，就是一个 DNA 螺旋的两条链中，只需要有一条链发生断裂，以暴露出 DNA 联会所需要的 DNA 链。人们认为这个在磷酸二酯骨架上发生的断裂，使得存在切口的链末端中的一个与其碱基配对的互补链分离，并使它能够自由地与另一个完整的 DNA 螺旋形成一个短的异源双链连接，从而开始了联会过程。这种模型在理论上是合理的，因此将近 30 年内教科书都这样描述。

在 19 世纪 90 年代早期，出现了许多灵敏的生物化学技术，可以用来确定在有丝分裂不同时期酵母染色体中形成的重组中间物的实际结构。这些研究表明，一般性重组由一种特殊的内切核酸酶起动，它能同时切断双螺旋的两条链，使 DNA 分子完全断裂开。断裂处的 5′端由一种外切核酸酶向内酶切，产生一个突出的 3′端单链。正是这些单链找到与其配对的同源 DNA 螺旋，导致在父代和母代染色体之间形成“连接分子”（图 5-56）。

下一部分，我们开始解释 DNA 单链是如何能“找到”同源双链 DNA 分子，从而开始 DNA 联会的。

DNA 杂交为一般性重组中碱基配对步骤提供了一种简单模型

对于一般性重组的联会步骤中起中心作用的碱基配对相互作用过程，其最简单的形式可以通过在试管中使 DNA 由分开的两条单链重新形成双螺旋来进行模拟。这个过程称为 DNA 复性或者**杂交**（hybridization），当出现罕见的随机碰撞将两条配对的 DNA 单链上的互补核酸序列并排在一起的时候，就会发生杂交，从而在这两条单链间形成一小段双螺旋。这个相对较慢的螺旋成核步骤后面，接着就是迅速的“拉链”步骤，此时双螺旋区域延展开，使发生配对互相作用的碱基数目最大化（图 5-57）。

通过这种方式形成新的双螺旋要求发生退火的两条链处于开放的，未折叠的构象。由于这个原因，体外杂交反应都在高温或存在有机溶剂如甲酰胺的条件下进行；这些条件使一条单链内因自身回折，发生碱基配对互相作用形成的短发卡螺旋解链。大部分细胞在这种剧烈条件下都不能存活，因而换用单链 DNA 结合（SSB）蛋白将发卡螺旋解链，并帮助与它们的互补单链退火。该蛋白对于 DNA 复制（前面描述过）以及一般性重组都很重要；它牢固而且协同地结合到 DNA 上所有单链 DNA 区域的糖-磷酸骨架上，使它们保持碱基裸露的延展构型（图 5-17 和图 5-18）。在这种伸展构型中，一条 DNA 单链可以有效地和一条核苷三磷酸分子（DNA 复制中）或者另一条 DNA 单链中互补部分进行配对（作为遗传重组过程的一部分）。

一般性重组的联会步骤中，一条 DNA 单链要找到的伴侣分子是一条 DNA 双螺旋，而不是另一条单链 DNA（图 5-56）。下一部分，我们来看重组过程中细胞是如何完成允许 DNA 进行杂交的关键事件的，一条单链 DNA 开始侵入到 DNA 双螺旋中。

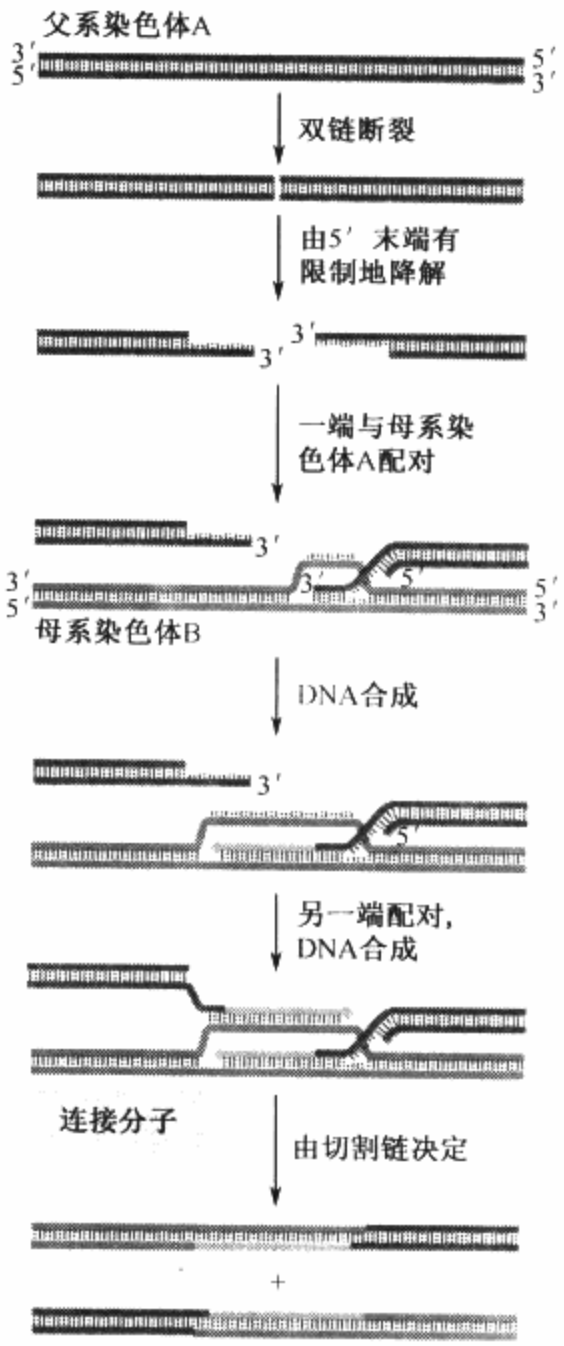


图 5-56　减数分裂中的一般性重组

如图所示，当一个内切核酸酶在染色体中产生双链断裂时，就会开始此过程。然后外切核酸酶产生两个突出的 3′单链末端，由它们找到另一条染色体上的同源区域并开始 DNA 联会。形成的连接分子最终能由选择性链切割拆分，产生两个已经发生交换的染色体，如图所示

RecA 蛋白及其同源蛋白质使 DNA 的一条单链能够与 DNA 双螺旋的同源区域配对

一般性同源重组比刚才描述的有单链 DNA 参与的简单杂交反应更加复杂，它需要几种特殊蛋白质。尤其是大肠杆菌 RecA 蛋白在染色体间重组过程中发挥中心作用；它和酵母、小鼠以及人类中的同源蛋白质使联会成为可能（图 5-58）。

与单链 DNA 结合蛋白类似，这种 RecA 蛋白牢固地而且以一长串的形式协同地结合到单链 DNA 上，形成核蛋白纤丝。由于每个 RecA 单体不止一个 DNA 结合位点，所以一个RecA

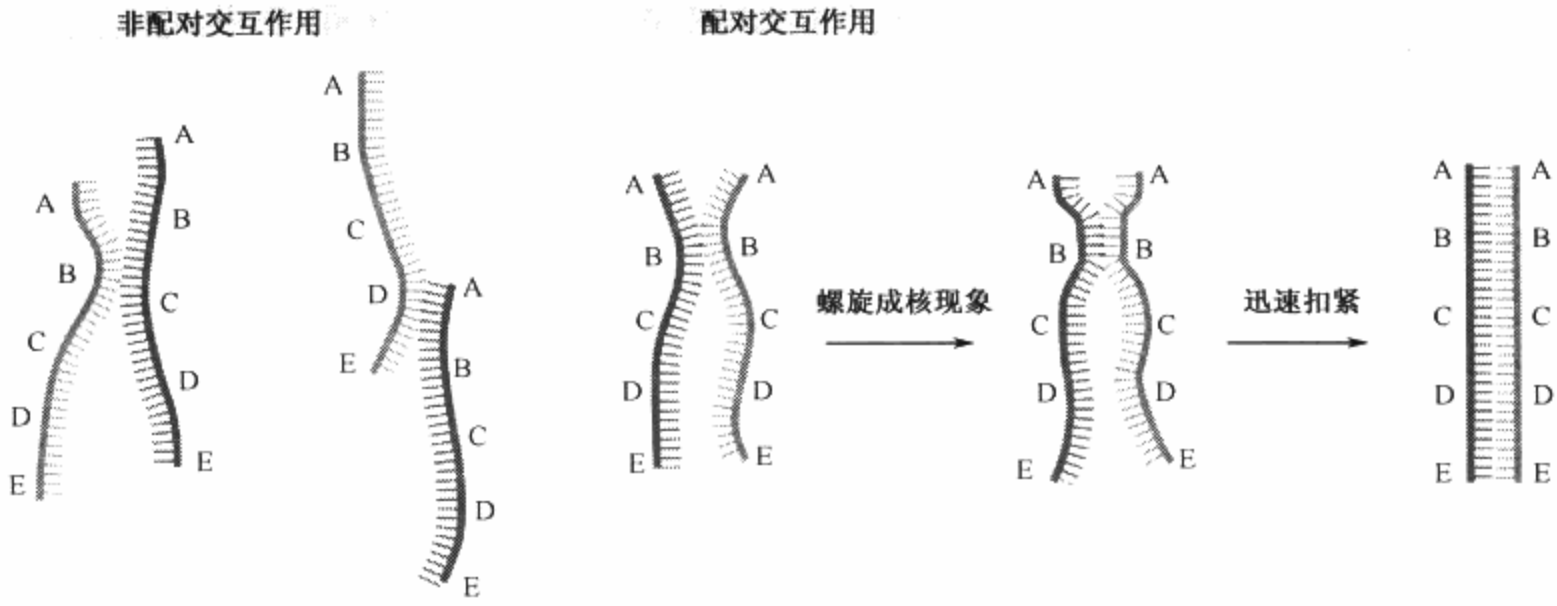


图 5-57　DNA 杂交

在依赖于两条互补 DNA 链随机碰撞的反应中，DNA 双螺旋能从两条分开的链重新形成。大多数这种碰撞都不会有什么结果，如左图所示，但是也有一些导致形成互补碱基配对的短区域（螺旋成核作用）。接着迅速的拉链反应导致形成完整的双螺旋。通过这种“尝试与错误”法，即使在数百万不配对的 DNA 链中，一条 DNA 链也会找到与其互补的链。一种相关的，高效的互补链 DNA 序列“尝试与错误”识别似乎起始了所有的一般性重组事件

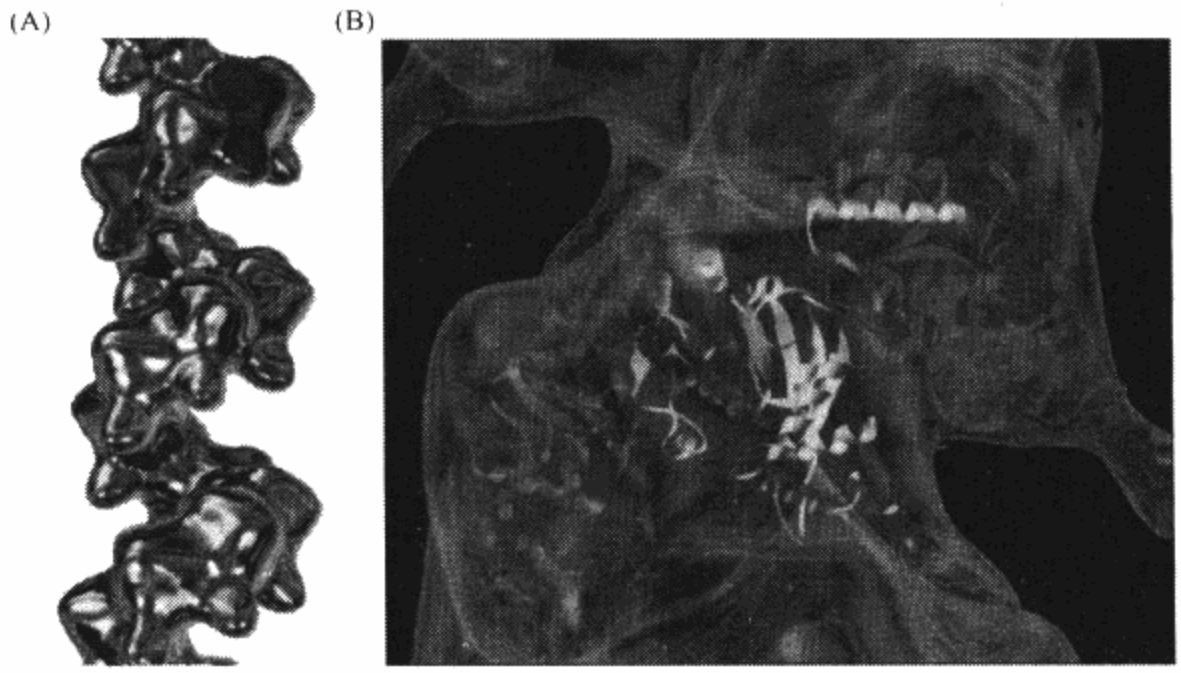


图 5-58　RecA 和 Rad51 蛋白-DNA 纤丝的结构

(A) 结合 DNA 单链的 Rad51 蛋白。Rad51 是细菌 RecA 蛋白的人源同源蛋白质；该螺旋纤丝中有 3 个连续的单体着以彩色。(B) 一小部分的 RecA 纤丝，该蛋白质的三维结构与电镜显微图片恰好吻合。这两个 DNA-蛋白质纤丝看起来很相似。每一圈螺旋大约有 6 个 RecA 单体占据该蛋白质展开的单链 DNA 中 18 个核苷酸。该结构中 DNA 确切的路径还不清楚[(A)经 Edward Egelnman 允许后使用；(B) 引自 X. Yu et al., *J. Mol. Biol.* 283：985～992，1998]

纤丝可以将一条单链和一个双螺旋维系在一起。这使其能够催化 DNA 双螺旋和单链 DNA 同源区域之间的多步 DNA 联会反应。通过一个三链中间产物，其中 DNA 单链与双链 DNA 分子大沟中的螺旋上暴露出来的碱基形成短暂的碱基配对，使得双链 DNA 打开之前，就已确定出同源区域（图 5-59）。这个反应起动了前面图 5-56 中显示的配对过程，因而它起动了两个重组 DNA 双螺旋之间链的交换。

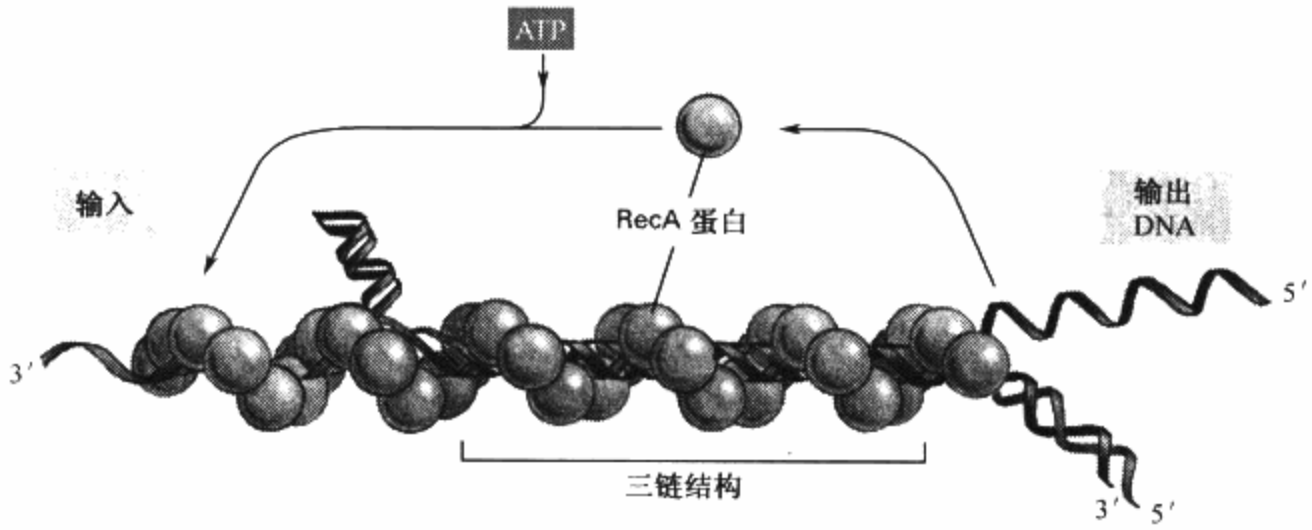


图 5-59　RecA 蛋白催化的 DNA 联会

体外实验显示，RecA 蛋白（红色）覆盖的 DNA 单链和 DNA 双螺旋（绿色）之间形成几种类型的复合体。首先形成的一个非碱基配对的复合物，一旦发现同源区域，它通过短暂的碱基翻转（图 5-51）转变为一个三链结构。该复合物不稳定，因为它需要一种异常的 DNA 形式，并且它延长了 DNA 异源数量（一条链绿色，一条链红色），再加上一条从原先的螺旋被取代的单链（绿色）。因此，图中所示的结构向左迁移，缠绕“输入 DNA”同时，产生“输出 DNA”，净结果是产生与图 5-56 中所画的相同的 DNA 链交换（摘自 S. C. West，*Annu. Rev. Biochem.* 61：603～640，1992）

一旦发生了 DNA 联会，来自两条不同 DNA 分子的链开始配对的短异源双链区域就通过一个称作分支移动的过程扩大。在任何具有相同序列的两条单链 DNA 要与互补的两条相同的互补链配对的地方，都能发生分支移动；在此反应过程中，一条单链上未配对的区域取代了另一条单链上已配对的区域，分支点向前移动的同时却不改变 DNA 碱基对的总数目。尽管能够发生自发的分支移动，但是它向两个方向同时前进，所以基本上没有任何变化，也不可能有效地完成重组过程[图5-60(A)]。RecA 蛋白催化单方向的分支移动，容易形成一个长达几千个碱基对的异源双链 DNA 区域[图5-60(B)]。

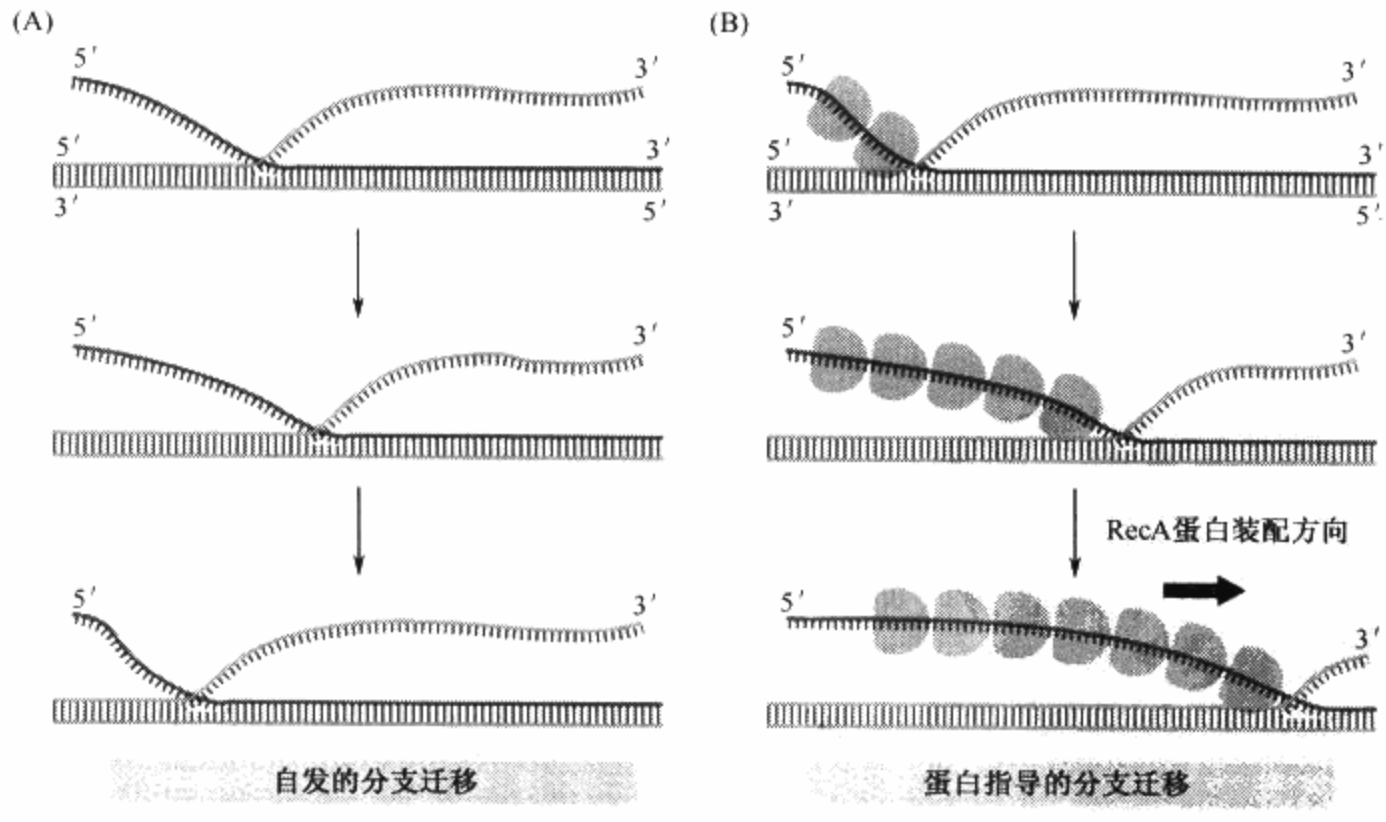


图 5-60　体外实验中观察到的两种分支迁移类型

（A）自发的分支迁移是一个来来回回，随机走动的过程，因此它在长距离中几乎不移动。（B）RecA 蛋白指导的分支迁移以已知的速率向同一方向前进，并可能由 RecA 蛋白纤丝在 DNA 单链上的极化装配推动，这个观察以图中所示的方向进行。甚至更加有效地催化分支迁移的特殊 DNA 解旋酶也参与了重组（图 5-63）

催化分支向一个方向移动有赖于 RecA 蛋白进一步的功能。除了具有两个 DNA 结合位点之外，RecA蛋白也是一个依赖 DNA 的 ATP 酶，它有另外一个位点用于结合并水解 ATP。这个蛋白质结合 ATP 时比结合 ADP 时与 DNA 结合得更牢固。此外，结合有 ATP 的新 RecA分子优先加入到RecA蛋白纤丝末端，然后 ATP 水解成 ADP。因此，DNA 上形成的 RecA 蛋白纤丝可能与肌动蛋白微管构成的细胞骨架纤丝一样，具有某些共同的动力学特征（讨论见第 16 章）。例如，该蛋白质能向一个方向沿一条 DNA 链“稳态活动”的能力可以驱动图 5-60（B）中显示的分支移动反应。

真核生物中存在 RecA 蛋白的多种同源蛋白质，它们每一种都特化后具有特异的功能

将真核生物中催化基本遗传功能的蛋白质与细菌（如大肠杆菌）中的类似蛋白质比较以后，通常会发现存在进化上相关的蛋白质，它们催化相似的反应。然而，在许多情况下，多种真核生物的同源蛋白质取代了特定的细菌蛋白质，它们特化后具有细菌蛋白质某一方面的特殊功能。

这种结论也适用于大肠杆菌的 RecA 蛋白：人类和小鼠至少有 7 种 RecA 同源蛋白质。推测每一种同源蛋白质都具有一种特殊的催化活性，以及自身的一套辅助蛋白质。Rad51 蛋白是酵母、小鼠以及人类中一种特别重要的 RecA 蛋白；在体外实验中，它催化一条单链 DNA 和一条 DNA 双螺旋之间的联会反应。对 Rad51 蛋白进行突变的遗传研究表明，这个蛋白质对上述三种生物都很重要，因为需要它来修复每个正常的 S 期中意外断裂的复制叉。这个蛋白质正确地发挥功能需要多种辅助蛋白质。其中 2 个，Brca1 和 Brca2 蛋白，首先被发现，因为它们基因发生的突变能够在人类后代继承下来，使它们患乳腺癌的频率大大增加。去掉 Rad51 蛋白会杀死细胞，而这样一种辅助蛋白质的改变能导致 Rad51 功能发生较为温和的改变，这些改变被认为导致 DNA 损伤的积累，DNA 损伤的积累在一小部分的细胞中经常产生癌（图 23-11）。

真核生物中不同的 RecA 同源蛋白专门用于减数分裂或者其他了解较少的类型独特的 DNA联会事件。有可能只有当特殊的 DNA 结构或细胞内条件允许该蛋白质结合到某个位点时，每一种真核生物 RecA 同源蛋白才会结合到一条 DNA 单链上，起动一般性重组的过程。

一般性重组常常涉及一个 Holliday 连接的形成

在两条不同双螺旋之间交换第一条单链的联会过程推测可能是一般性重组中缓慢且困难的步骤（图 5-56）。此步以后，延伸配对区域以及在两个 DNA 螺旋之间建立进一步的链交换就会迅速地进行。在大多数情况下，结果形成一个重要的重组中间物，Holliday 连接（也称为交叉链互换）。

在 Holliday 连接（Holliday juncton）中，两条最初配对的同源 DNA 螺旋通过相互交换 4 条链中的两条得以维系在一起，每一个螺旋提供一条链。如图 5-61（A）所示，可以认为 Holliday 连接含有两对链：一对交叉链和一对非交叉链。然而，通过特殊蛋白质催化的一系列旋转运动，这个结构能够发生异构化，形成一个更加开放的结构，其中两对链占据相等的位置[图5-61(B)和图5-62]。反过来，这个结构也能异构化，形成很像最初连接的构象，除了交叉链已转变成非交叉链以外，反之亦然[图5-61(C)]。

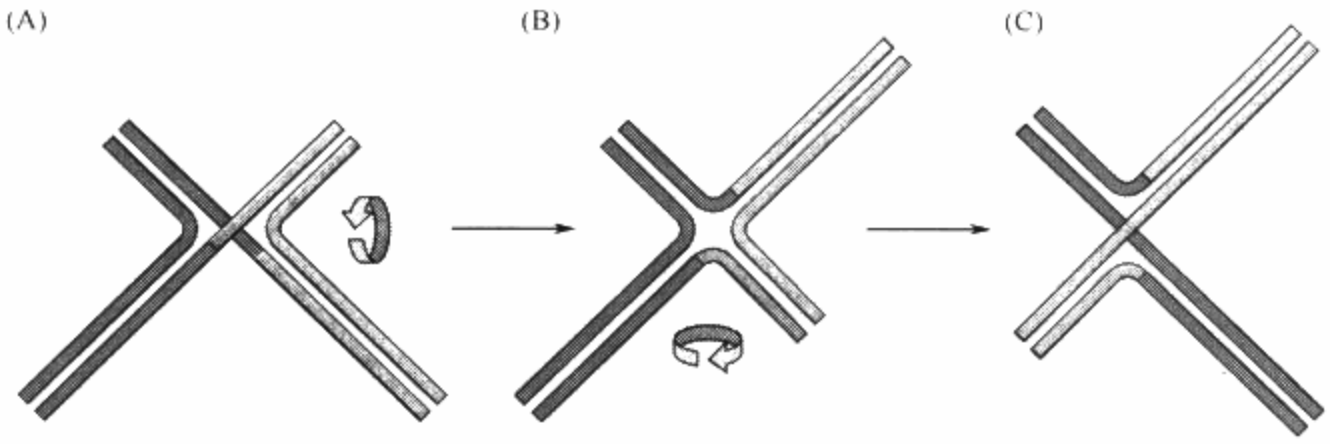


图 5-61　Holliday 连接及其异构化

正如文中描述的那样，一般性重组中的联会步骤是由结合 DNA 单链的 RecA 类型蛋白质催化，该步骤后面常常是由此彼此配对的两条 DNA 双螺旋之间发生链互换。这种交换形成一种称为 Holliday 连接的独特 DNA 结构，该结构以首先提出它的形成的科学家命名。（A）最初形成的结构含有两条交叉（内部）和两条未交叉（外部）的链。（B）Holliday 连接的异构化产生一个开放的对称结构。（C）进一步的异构化能互相转换交叉和未交叉链，形成一个在其他情况下与（A）相同的结构

一旦 Hollidy 连接形成了一个开放的结构，一系列的特殊蛋白质就会加入到连接中来：这些蛋白质中有一种使用 ATP 水解释放的能量沿两个螺旋迅速地移动交叉点（两个 DNA 螺旋发生连接的位点），将异源双链 DNA 区域向远处延伸（图 5-63）。

为了重新形成两个分开的 DNA 螺旋，并且因此结束交换过程，Holliday 连接中连接两个螺旋的链必须最终切断，整个过程称作拆分。有两种方可以拆分 Holliday 连接。在其中一种方式中，最初的一对交叉链被切断[入侵或内侧链，见图5-61(A)]。在这种情况下，两条最初的 DNA 螺旋彼此分离，几乎不发生变化，交换了形成异源双链的单链 DNA。在另一种方式中，最初的一对非交叉链被切断[图5-61(C)中的内侧链]。现在结果将深远得多：形成两条重组染色体，通过交换事件相互交换了双链 DNA 的大部分片段（图 5-64）。

遗传分析表明，在重组过程中很容易形成长达几千碱基对的异源双链区域。下面将会描述，对这些异源双链（通常由几乎相同片段的互补链组成）的加工能进一步改变最终产生的 DNA 螺旋的信息。

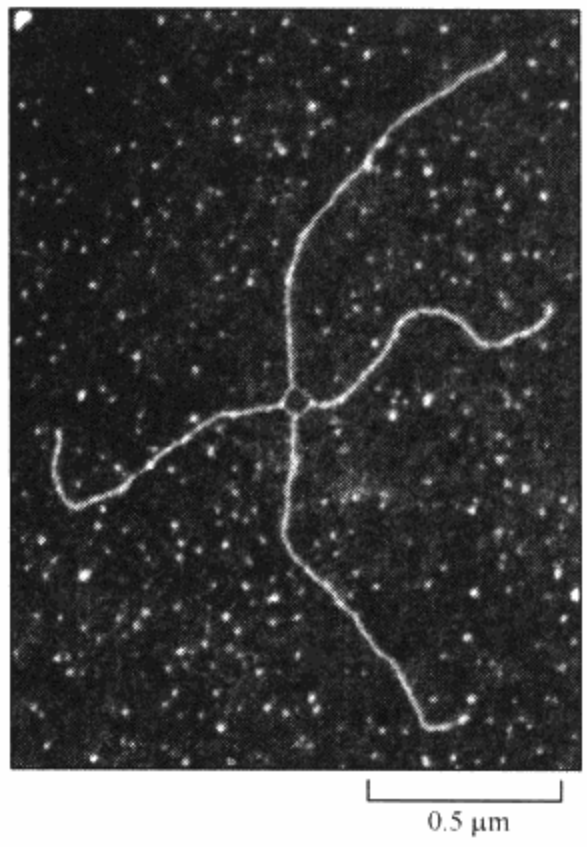


图 5-62　Holliday 连接的电子显微照片

这幅 Holliday 连接的图对应于图5-61(B)中显示的开放结构（经 Huntington Potter 和 David Dressler 允许后使用）

一般性重组能引起基因转换

在有性繁殖的生物体中，有一个基本的遗传规则，就是父母对后代做出相等的遗传贡献，而后代分别从父母那里继承一套完整的基因。因此，当二倍体细胞进行减数分裂生成 4 个单倍体细胞时（讨论见第 20 章），确切地讲应该是这些细胞内一半的基因是母源的（二倍体细

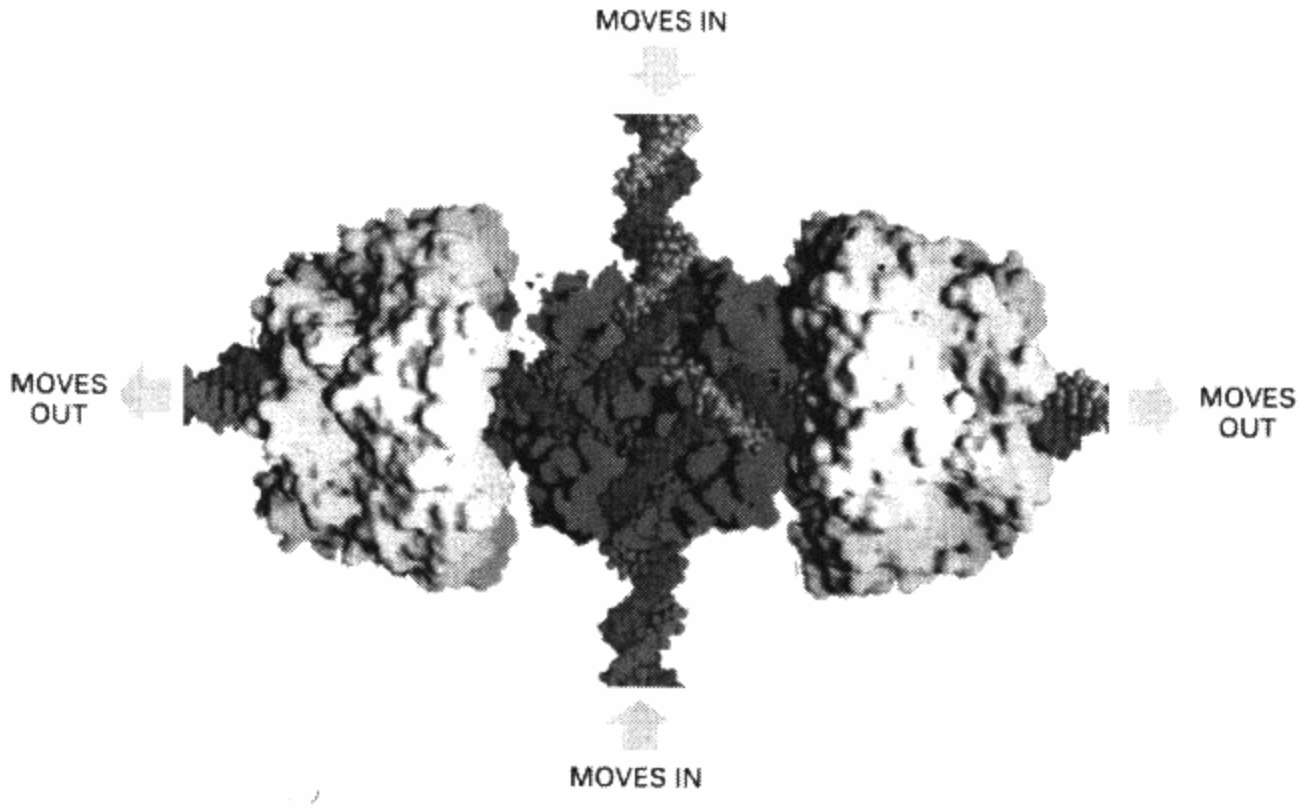


图 5-63　Holliday 连接处发生的酶催化的双分支移动

在大肠杆菌中，一个 RuvB 蛋白的四聚体（绿色）和两个 RuvB 蛋白的六聚体（淡灰色）结合到开放的连接形成。RuvB 蛋白利用 ATP 水解能量将交换位点沿配对的 DNA 螺旋迅速移动，如图所示的那样扩展异源双链区域。有证据表明类似的蛋白质也在脊椎动物细胞中行使该功能（图像经 P. Arty miuk 允许使用；修改后摘自 S. C. West，*Cell* 94：699～701，1998）

胞从它母亲继承的基因），另一半是父源的（二倍体细胞从它父亲继承的基因）。在有些生物体（如真菌）中，有可能获得并分析由单个细胞通过减数分裂产生的全部 4 个单倍体配子。在这类生物中进行的研究表明发生违反标准遗传法则的罕见情况。然而，例如，减数分裂产生了 3 份母源基因的拷贝，只有 1 份父源等位基因的拷贝［同一基因的其他版本称为等位基因（alleles）］。这种现象称为**基因转换**（gene conversion）（图 5-65）。减数分裂中，基因转换常常与同源遗传重组一同发生（在有丝分裂中较不常见），人们认为这是一般性重组与 DNA 修复的直接结果。遗传研究表明，只有一小部分 DNA 典型地发生基因转换，并且在许多情况下一个基因只有一部分发生了改变。

在基因转换过程中，DNA 序列信息从保持不变的一条 DNA 螺旋上（供体序列）转移到另一条序列发生改变的 DNA 螺旋上（受体序列）。这个过程以几种不同的方式进行，每一种方式都涉及以下的两个过程：①将两条同源 DNA 双螺旋并排的同源重组事件；②局部 DNA 的有限合成，它对于产生一个等位基因而另外一份拷贝是必需的。在最简单的例子中，一般性重组过程形成一个异源双链（图 5-55），其中配对的两条 DNA 链序列并不完全相同，而且含有一些错误的碱基。如果其中一条链上的错配核苷酸被催化错配修复的 DNA 修复酶识别并去除，那么就会在相反链上产生这条 DNA 序列额外的一份拷贝（图 5-66）。相同的基因转换过程不进行交换也能发生，因为基因转换只需要一条单链 DNA 侵入一个双螺旋，形成一个短的异源双链区域。人们认为后面这种基因转换类型能引起遗传信息异常容易地发生转移，而这种情况通常只在串联排列的重复基因不同拷贝之间才会观察到。

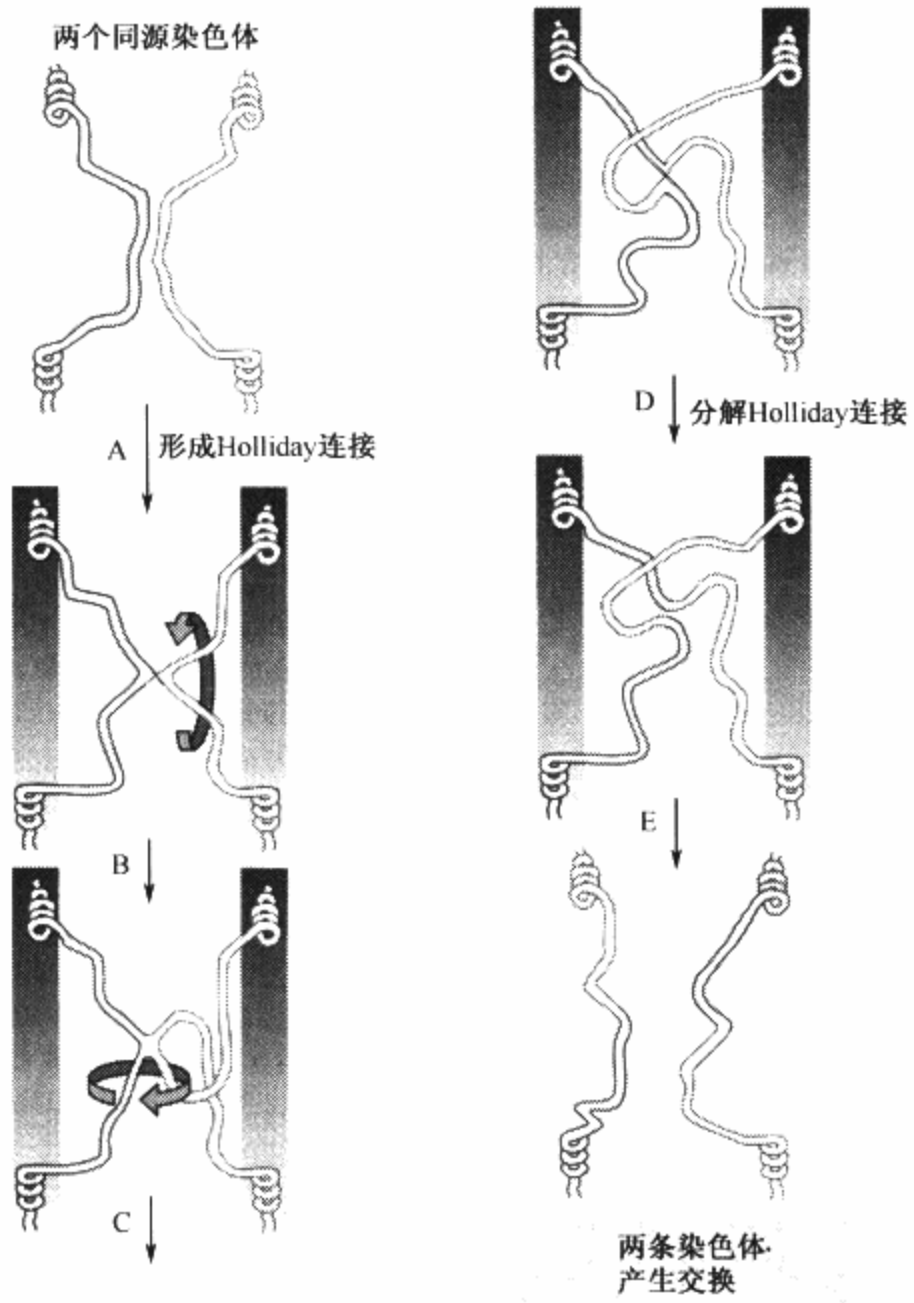


图 5-64　Holliday 连接的拆分产生交换的染色体

本例中，红色和绿色染色体的同源区域通过交换两条链已形成一个 Holliday 连接。将这两条链切开将终止交换，不发生互换。如果 Holliday 连接发生异构化（步骤 B 和 C），最初的非交叉链就成为两条交叉的链，现在将它们切开就产生两条交换的 DNA 分子（底部）。在有丝分裂的一般性重组过程中，这种异构化可能需要断裂和再连接两条同源 DNA 双螺旋。之所以画出中间嵌板上的灰色条带，是因为要清楚地表明图中所示的异构化不会干扰两条染色体其余部分

一般性重组事件在有丝分裂和减数分裂细胞中倾向于产生不同的结果

我们已经看到，减数分裂重组的起始具有鲜明的特征——发生重组的染色体中双螺旋的两条链都断开。那么随后进行的减数分裂过程与同样基于一般性重组且细胞用于精确地修复染色体中意外的双链断裂的机制之间如何区分呢（图 5-53 中的同源末端连接反应）？在两种情况下，由双链断裂产生的两个染色体新末端都发生降解过程，从而暴露出具有突出 3′端的一条单链。此外，在两种情况下，这条链都找到具有相同核苷酸序列的区域的完好 DNA 双螺旋并与之发生由 RecA 类蛋白质催化的聚合反应。

对于双链断裂的修复，DNA 合成利用受体 DNA 螺旋的一条链作模板，将 3′端延伸达数千个核苷酸。如果另一个断裂末端同样参与了联会反应，就会形成一个连接分子（图 5-56）。根据随后进行的反应的不同，最终结果或者是最初 DNA 螺旋经过双链断裂的修复得以恢复

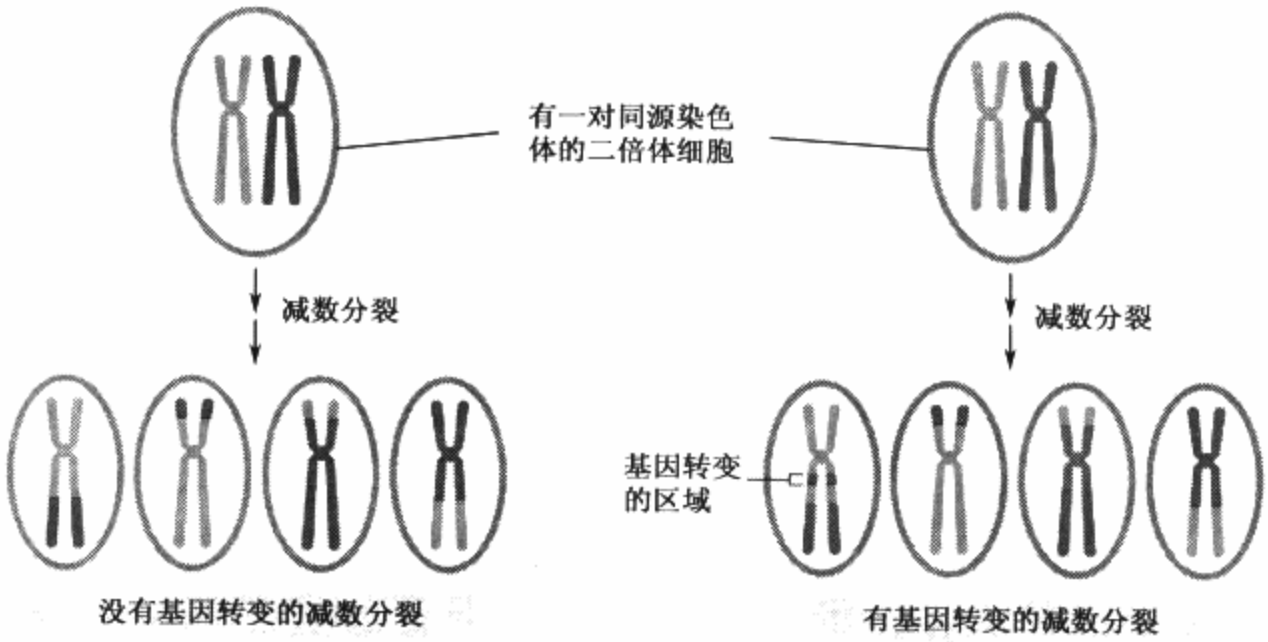


图 5-65　减数分裂中的基因转换

正如第 20 章描述的那样，减数分裂是一个二倍体细胞产生 4 个单倍体细胞的过程。生殖细胞（如卵子和精子）通过减数分裂形成

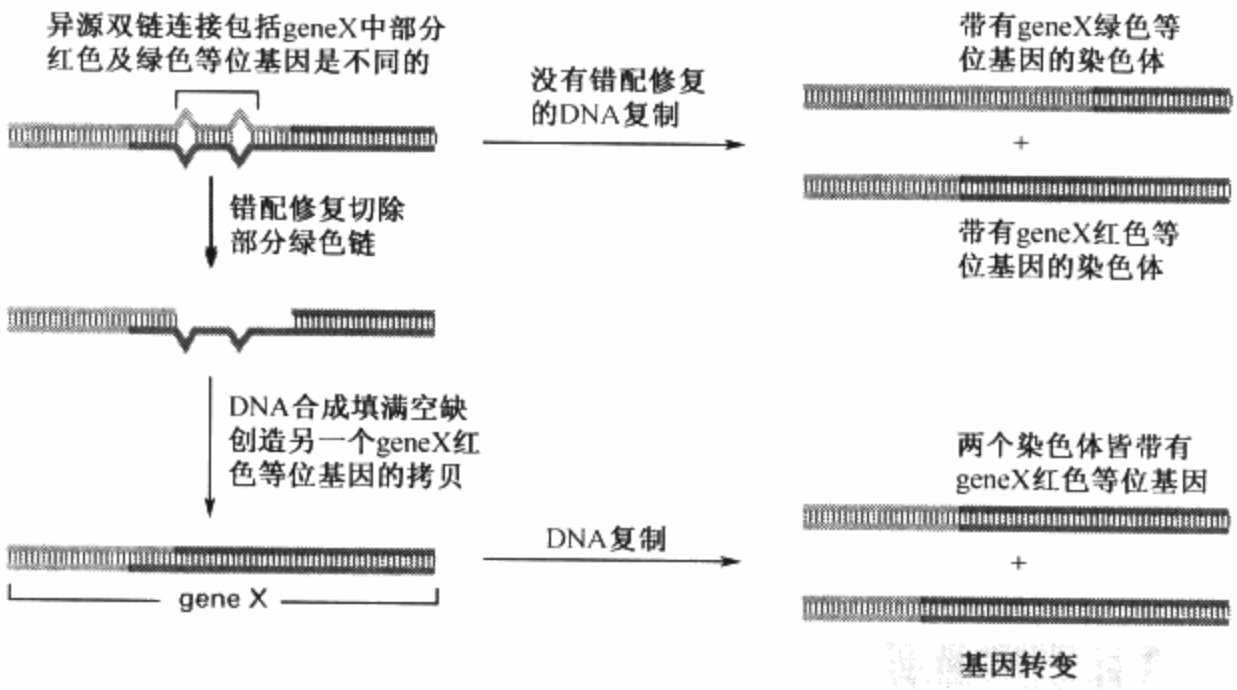


图 5-66　错配校正引起的基因转换

在这个过程中，异源双链连接在同源的父源和母源染色体交换区域形成。如果父源和母源 DNA 序列稍有不同，异源双链连接就会包含几个错配碱基对。结果在双螺旋中形成的错配然后可能由 DNA 错配修复机器（图 5-23）进行纠正。该机器能清除父源或母源链上的核苷酸。这个错配修复的结果，就是基因转换，发现它是通常在减数分裂发生的父源和母源等位基因同等拷贝分离的一种偏差情况

原状（在有丝分裂细胞中占大多数），或者是发生交换，使异源双链连接将两条不同的 DNA 螺旋维系在一起（在减数分裂细胞中占主要地位的反应）。人们认为是这些指导这些细胞进行有丝分裂细胞的一套特殊蛋白质引起交换事件的发生。这些蛋白质不仅确保了具有 2 个 Holliday 连接的连接分子的形成，而且引起 2 个连接处各自不同链的配对，因而导致了交换的发生（图 5-67）。

无论一般性重组发生哪种结果，所涉及的 DNA 合成在双链断裂位点将一些遗传信息转换到了另一条同源染色体上对应位置。如果这些区域代表同一基因的不同等位基因，在断裂

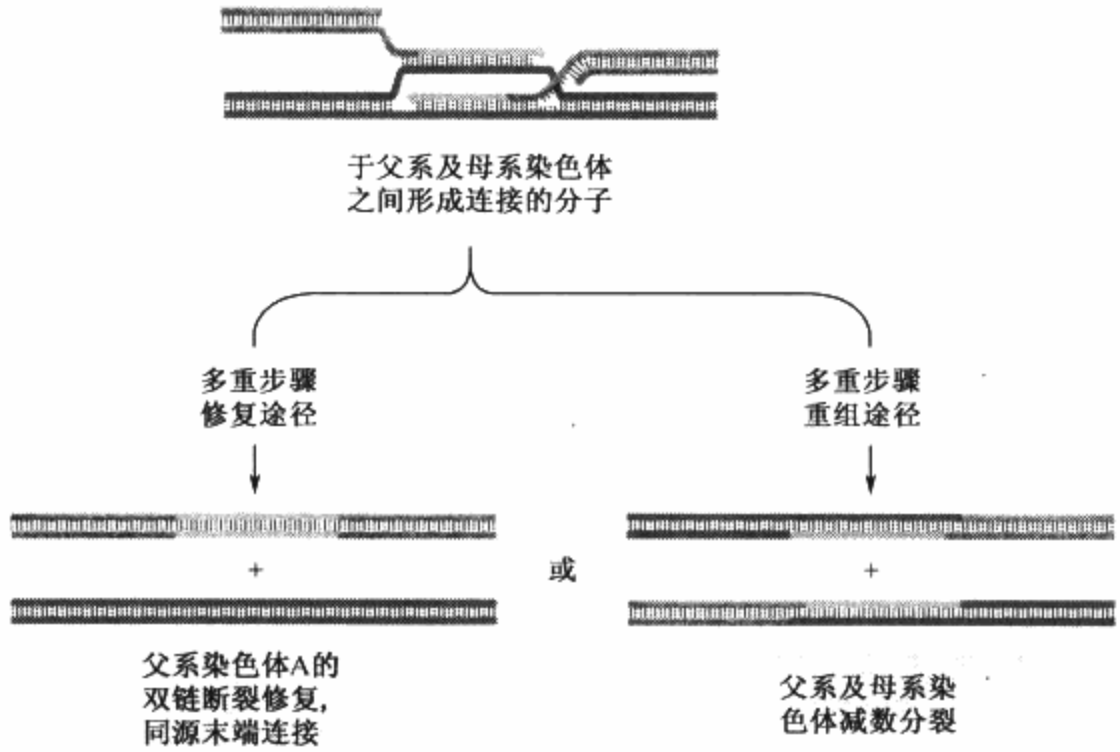


图 5-67　在有丝分裂和减数分裂中一般性重组中间物的不同拆分

如前面的图 5-56 所示，当一个双螺旋（绿色）中产生一个双链断裂时，就开始一般性重组过程，然后开始 DNA 降解和链侵入同源的 DNA 双链（红色），随后合成新的 DNA（橙色），以产生图中所示的连接分子。根据后面发生事件的不同，连接分子的拆分或者导致原先的双链断裂的精确修复（左边）或者引起染色体交换（右边）。利用实验方法在特异的 DNA 位点诱导双链断裂可以对有丝分裂和减数分裂细胞中一般性重组的结果进行量化。这些事件中超过 99%都不能在有丝分裂细胞中产生交换，而在减数分裂细胞中常常能产生交换结果。在任何一种情况下，如果母源和父源染色体 DNA 序列只是在这里所示的新 DNA 合成区域不同，那么在新 DNA 合成区域的绿色 DNA 双链区域转换成红色双链序列（基因转换）

螺旋中的核酸序列就会转换到完好螺旋的相同区域，引起基因转换。酿酒酵母利用伴随双链断裂修复的基因转换从一种交配类型转换成另一种交配类型（讨论见第 7 章）。在这个例子中，由一种 HO 内切核酸酶在酵母交配类型位点通过切断一段特殊 DNA 序列，有目的地引发了双链的断裂。在 DNA 降解将断裂位点的旧序列去除后，具有相反交配类型（a 或 α）的“交配类型盒” DNA 序列的断裂末端发生联会，并随之以前面提到的封闭染色体断裂区域的方式进行 DNA 合成，使得丢失的遗传信息得以恢复。实际上是经过对这种位置精确固定的双链断裂修复形式进行详细的研究之后，才揭示出同源末端连接的一般机制。

错配校对防止两条不能配对的 DNA 序列之间发生任意重组

前面讨论过，重组过程中关键的一步发生在两条互补 DNA 序列配对，在两个双螺旋之间形成一个异源双链连接的时候。用纯化的 RecA 蛋白进行体外实验表明，即使两条 DNA 链的序列不完全互补时，也能有效地进行配对。例如，当平均每 5 个核苷酸中只有 4 个能形成碱基对时。如果从这些错配的序列继续进行重组，将在细胞内尤其在基因组中含有密切相关 DNA 序列的细胞内，引起巨大的混乱。细胞是如何防止在这些序列间进行交换的呢?

虽然还没有完整的答案，但是在细胞和酵母中进行的研究表明，去除复制错误的相同错配校对系统的组分（图 5-23）也具有其他的功能，能够阻止不完全错配 DNA 序列间形成遗传重组。例如，长期以来人们知道在 2 个紧密相关的细菌，大肠杆菌和沙门菌中存在的同源基因通常不会发生重组，即使它们的核酸序列相似程度达到 80%。然而，一旦突变引起错配校对系统失活，这种物种间重组事件的发生频率将会增加 1000 倍。人们认为错配校对系统通

常在最初的链交换中识别错配碱基，并且，如果存在大量的错配，会阻止随后需要打断和重新连接 2 个配对 DNA 螺旋的步骤，这种机制保护了细菌基因组不能发生改变，否则就会因偶尔进入细胞的外来 DNA 分子与之发生重组而改变。在脊椎动物细胞中，含有许多紧密相关的 DNA 序列，人们认为存在相同的重组校对帮助阻止发生随意的重组，否则将会搅乱基因组（图 5-68）。

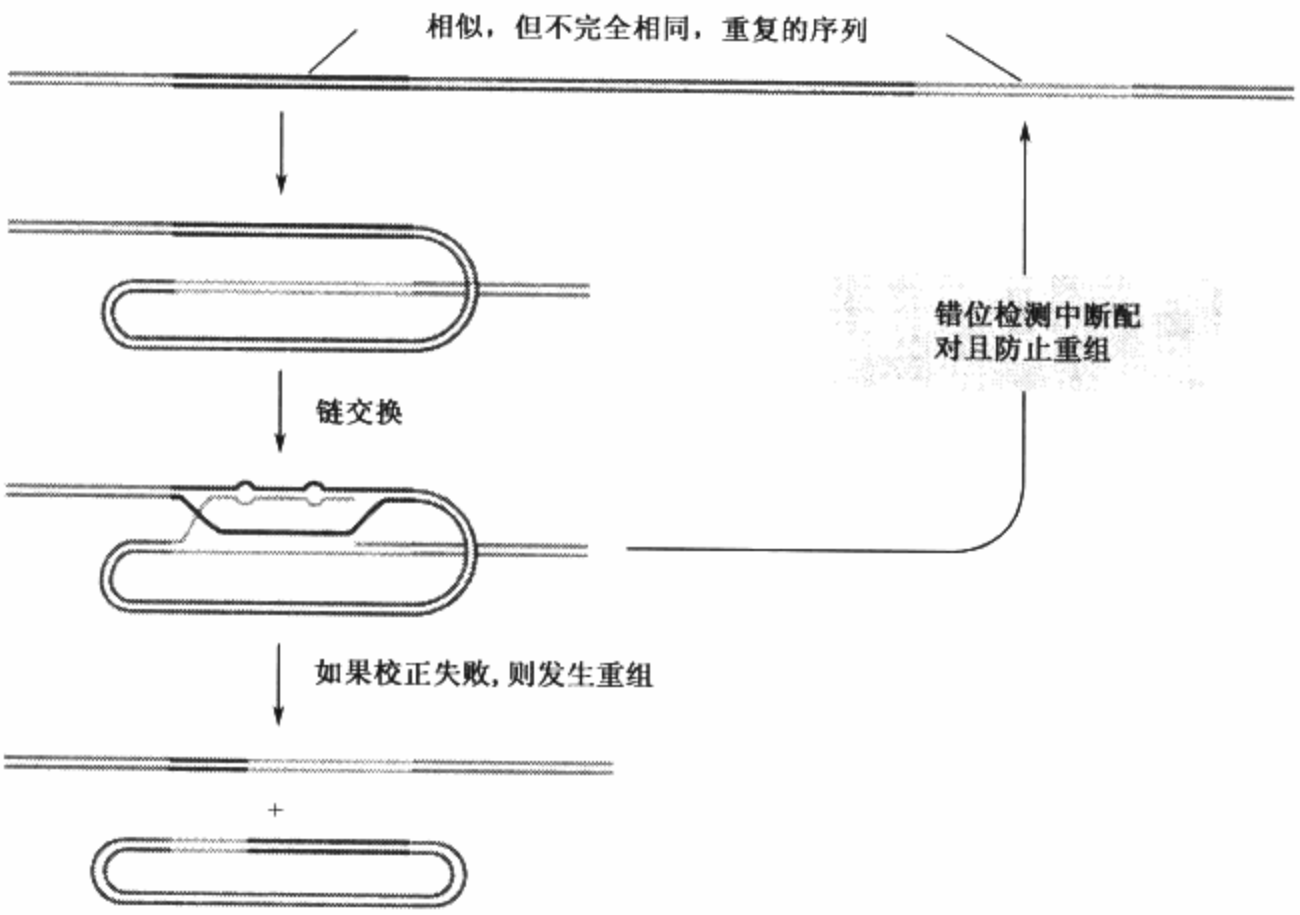


图 5-68　防止一般性重组破坏含有重复序列的基因组的稳定的机制

对细菌和酵母细胞研究后表明前面图 5-23 中图解的错配校对系统组分还具有该图所示的另外功能

小结

一般性重组（也称作同源重组）使 DNA 双螺旋的大段区域能够从一条染色体转移至另一条染色体，在真菌、动物和植物中，它引起减数分裂期间染色体发生交换。在所有细胞中，一般性重组对于维持细胞都是关键的，它常常起始于一个双链的断裂，随后经加工，裸露出一个单链 DNA 末端。这个单链与 DNA 双螺旋一段同源区域的联会由细菌 RecA 蛋白进行催化，联会常常导致形成一个含有 4 条链的结构，称为 Holliday 连接。根据将这个连接分离成 2 个独立双螺旋的链断裂模式的不同，产物可以是一个经过精确修复的双链断裂，也可以是发生交换的两条染色体。

由于一般性重组依靠发生重组的 2 个 DNA 双螺旋链间广泛的配对相互作用，所有它只在同源 DNA 分子间进行。基因转换，即基因信息从一条染色体向另一条染色体的非对等转移，是一般性重组的结果，其间也涉及有限数量相关 DNA 的合成。

位点特异性重组

在一般性重组中，序列很相似的 DNA 片段之间发生了 DNA 重排。尽管这些重排能导致染色体间等位基因的交换，但是相互作用的染色体上基因顺序常常保持不变。另一种类型的

重组，**位点特异性重组**（site-specific recombination）能改变基因的顺序，而且还能向基因组添加新信息。位点特异性重组在基因组内非同源位点之间转移特化的核酸序列，这些核酸序列称作可移动遗传元件。这种转移既可以在同一染色体不同位置之间发生，也可以在两条不同染色体之间发生。

可移动遗传元件的大小在几百个到几万个核苷酸对之间变化，实际上在所有检验过的细胞中都能确定它的存在。这些元件中有一些是病毒，其中的位点特异性重组用于将它们的基因组移入或移出它们的宿主细胞。病毒可以将它的核酸包装到病毒颗粒中，而病毒颗粒可以通过细胞外环境从一个细胞转移到另一个细胞中。许多其他的可移动元件只能在单个细胞内移动（以及它的后代细胞），缺乏离开它的寄主细胞的任何内在能力。

位点特异性重组事件的遗迹可以组成基因组可观的一部分。在许多脊椎动物中发现的大量重复 DNA 序列大多来自可移动遗传元件；实际上，这些序列占人类基因组的至少 45%（图 4-17）。长期以来，这些元件的核苷酸序列已经因随机突变发生了改变。因此，我们 DNA 中这些元件的许多拷贝只有一部分仍然保持活性并能移动。

除了移动它们自身以外，所有类型的可移动遗传元件偶尔也能转移或重排宿主细胞基因组中相邻 DNA 序列。例如，这些移动可以引起邻近核苷酸序列的删除，或者能将这些序列携带至其他位点。通过这种方式，与一般性重组一样，位点特异性重组产生许多进化所依赖的基因改变。可移动元件的易位在大量不同的生物体，包括人类中，引起自发突变；其中在有些生物（如果蝇）中，我们已经知道观察到的基因中大部分是因这些元件产生的。因此，长期以来位点特异性重组导致在基因组中发生了大部分的重要的进化改变。

可移动遗传元件能够通过转座或者保守机制进行移动

与一般性重组不同的是，位点特异性重组由识别短且特异的核苷酸序列的重组酶指导进行，这些序列可以存在于发生重组的 DNA 分子的一条中，也可以在两条中都存在。这种重组并不要求存在广泛的 DNA 同源性。每种类型的移动元件通常编码介导自身转移的酶，并含有该酶在其上发挥作用的特定位点。许多元件也携带其他基因。例如，病毒编码了能使它在细胞外生存的外壳蛋白，以及重要的病毒蛋白酶。携带抗生素抗性基因的移动元件的扩散是细菌群落中抗生素抗性大范围散播的主要原因（图 5-69）。

位点特异性重组可以通过两种不同机制进行，其中任何一种都需要特殊的重组酶和特异的 DNA 位点。①**转座的位点特异性重组**（transpositional site-specific recombination）通常需要位于染色体中的可移动 DNA 片段末端的断裂反应，以及将这些末端附着到许多不同的非同源目标 DNA 位点上。该过程不需要异源双链 DNA 的形成。②**保守的位点特异性重组**（conservative site-specific recombination）需要一个短异源双链连接的形成，因此它需要在供体和受体 DNA 分子上都相同的一段 DNA 序列。我们首先讨论转座的位点特异性重组（简称转座），然后在本章最后再讨论保守的位点特异性重组。

转座的位点特异性重组能将可移动遗传元件插入到任何 DNA 序列中

转座子（transposon），也称为**转座元件**（transposable element），是指那些通常只有适度的目标位点选择性的可移动遗传元件，因而能将它们自身插入到许多不同的 DNA 位点。在转座过程中，有一个特殊的酶，通常由转座子编码，因而称为转座酶，它作用于转座子 2 个末端的特殊 DNA 序列，首先将转座子从侧翼 DNA 上分离下来，然后将它插入新的目标 DNA 位点。该元件末端与插入位点末端并不要求具有同源性。

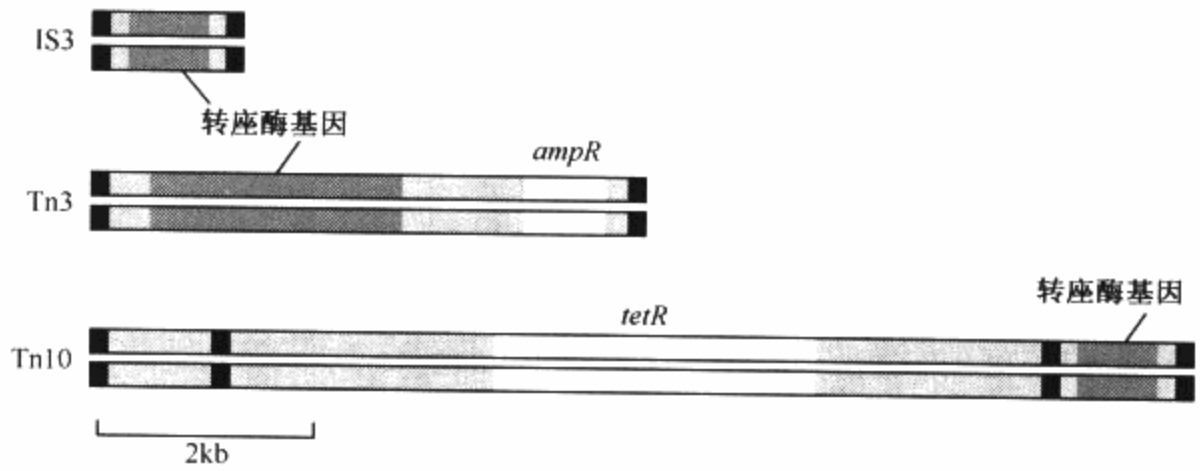


图 5-69　细菌中发现的许多种可移动遗传元件中的三种

这些 DNA 元件中每个都含有一个编码转座酶的基因，转座酶是进行元件移动所需的某些断裂和连接反应的酶。每个可移动元件也携带了短的 DNA 序列（红色表示），这些短 DNA 序列只由该元件编码的转座酶识别，并且对该元件的移动是必需的。此外，图中显示的三种可移动元件中有两种携带编码失活抗生素氨苄青霉素（*ampR*）和四环素（*tetR*）的抗性基因。底图所示的转座元件 Tn10 被认为是从将两个短的可移动元件偶然加到四环素抗性基因的任一边进化而来；四环素作为抗生素的广泛使用帮助了该基因在细菌群落中传播。图中所示的三种转座子元件全部都是 DNA 转座子（见正文）

大部分转座子仅仅在很稀少的情况下才会转移（对于细菌中的许多元件来说，10^5 个细胞世代中只发生一次），因为如此，通常难以将它们和染色体的非移动部分区分开。在大多数情况下，是什么引发了它们的转移还不为我们所知。

根据转座子的结构和转座机制不同，可将它们分为三大类（表 5-3），其中每一类在下面

表 5-3　三大类的转座元件

分类描述和结构	全部元件中的基因	移动模式	例子
仅限 DNA 的转座子 于两端的短、倒置的重复序列	编码转位酶	像 DNA 移动，不是切除就是遵循复制的途径	P 元件(果蝇) Ac-Ds(玉米) Tn3 和 IS1（大肠杆菌） Tam3（金鱼草）
像反转录病毒的反转录转座子 于末端直接地重复的长末端重复(LTRs)	编码反转录酶及相似反转录病毒	籍由 LTR 启动子制造的 RNA 中间物移动	Copia（果蝇） Ty1（酵母） THE-1(人类) Bs1(玉米)
非转录病毒的反转录转座子 多聚腺苷酸(Poly A)于 RNA 转录本的 3′末端；5′末端通常被截短	编码反转录酶	籍由通常由邻近的启动子制造的 RNA 中间物移动	F element(果蝇) L1(人类) Cin4(玉米)

注：这些元件长度由 1000 至 12 000 碱基对，每个家族含有许多成员，这里仅列出少数；除了转座的元件，有些病毒可以移进/出宿主细胞的染色体，这些病毒与前两类约，转座子有关联。

的部分中都将详细讨论。事实上，前面两类转座子使用了相同的 DNA 断裂和 DNA 连接反应进行转座。然而，对于 DNA 转座子而言，可移动元件在其整个生命周期都以 DNA 存在：发生转座的 DNA 片段直接从供体 DNA 切下，然后由转座酶连接到目标位点上。相比之下，逆转录病毒样转座子通过较为间接的机制转移。一种 RNA 聚合酶首先将可移动元件的 DNA 序列转录成 RNA，然后该酶使用合成的 RNA 分子作为模板将 RNA 分子逆转录成 DNA，最后正是这个 DNA 拷贝插入到基因组中新的位点。由于历史的原因，将催化这种插入反应的转座酶样的酶称为整合酶，而不是转座酶。表 5-3 中第三类转座子也是通过将 DNA 分子转录成 RNA 分子后再合成它的 DNA 拷贝后进行转移。然而这些非逆病毒的逆转座子利用的机制与前面刚描述的 RNA 分子直接参与转座反应的机制是不同的。

DNA 转座子提供 DNA 断裂和连接机制进行转移

许多 **DNA 转座子**（DNA-only transposon）通过**剪切-粘贴转座**（cut-and-paste transposition），使用图 5-70 列举的机制，从 1 个供体位点转移到目标位点。转座酶的每个亚基识别元件末端的同一特异 DNA 序列，这些亚基两两结合形成 1 个二聚体转座酶，并产生 1 个 DNA 环，将该元件的 2 个末端并列在一起。然后转座酶在这个 DNA 环 2 个末端进行切割，从而裸露出元件的末端，并从原来的染色体完全去除（图 5-71）。为了完成这个反应，转座酶催化了该元件的 2 个 DNA 末端对目标 DNA 的直接攻击，并在它将该元件和目标 DNA 连接起来的时候，在目标分子中断开 2 个磷酸二酯键。

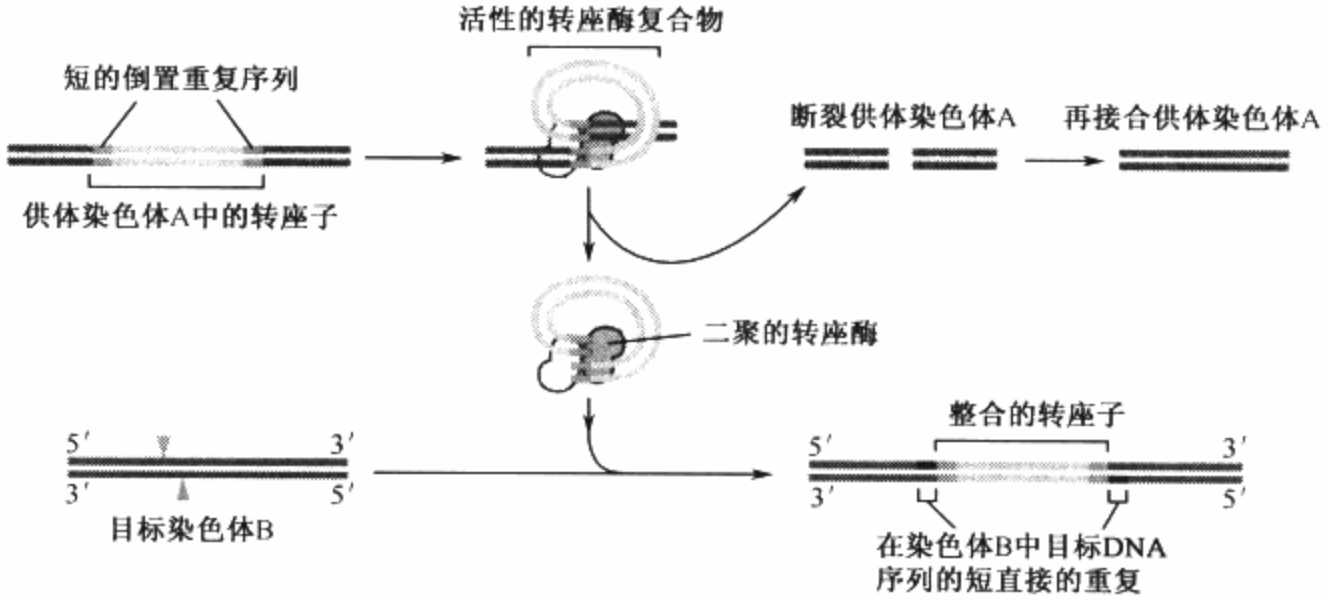


图 5-70　剪切和粘贴的转座

染色体中的 DNA 转座子可以由位于它们末端的“反向重复 DNA 序列”（红色）识别。实验研究表明这些可以短至 20 核苷酸的序列是对于它们之间的 DNA 被该元件相关的特定转座酶转座所必需的所有序列。DNA 转座元件经过剪切和粘贴过程，从染色体一个位点移动到另一个位点，当转座酶将两个反向 DNA 序列带到一起，形成一个 DNA 环之后才开始进行。由转座酶催化的插入目的染色体过程通过在目标染色体（红色箭头处）形成交错的断裂在随机位点发生。结果，如图所示，目标 DNA 序列短的直接重复标记插入位点。尽管供体染色体（绿色）中的断裂重新封闭，但断裂和修复过程常常改变 DNA 序列，引起被切割的转座元件的原先位点发生突变（未标出）

由于两条目标 DNA 链中形成的断裂是错开的（图 5-70 中的红色箭头），所以起初在产物 DNA 分子中形成了 2 个短的单链缺口，分别位于插入转座子的两端。这些缺口由宿主细胞 DNA 聚合酶和 DNA 连接酶进行填补，以便完成重组过程。期间产生相邻目标 DNA 序列的

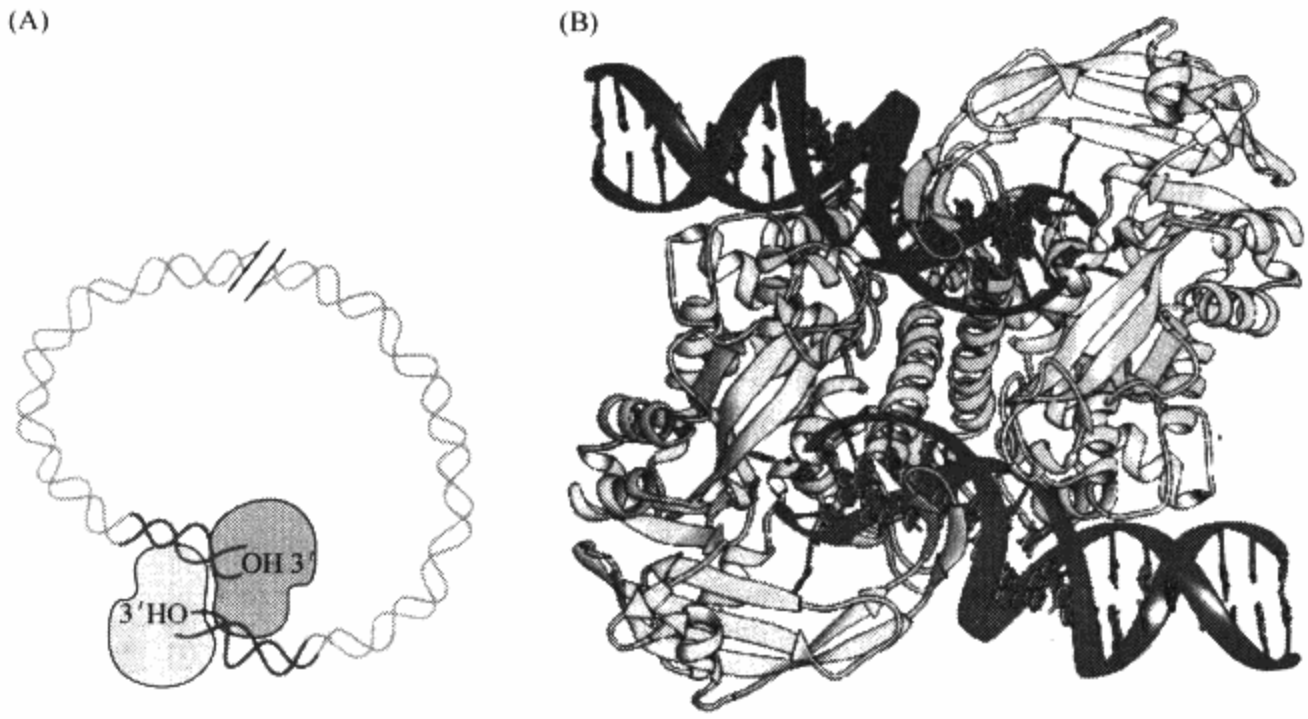


图 5-71　剪切和粘贴的转座酶形成的重要的中间产物的结构

（A）整体结构概图。（B）维系两个 DNA 末端的转座酶的详细结构，其 3′-OH 基团已作好攻击目标染色体的准备[(B)摘自 D. R. Davies el al.,*Science* 289：77～85，2000. c AAAS]

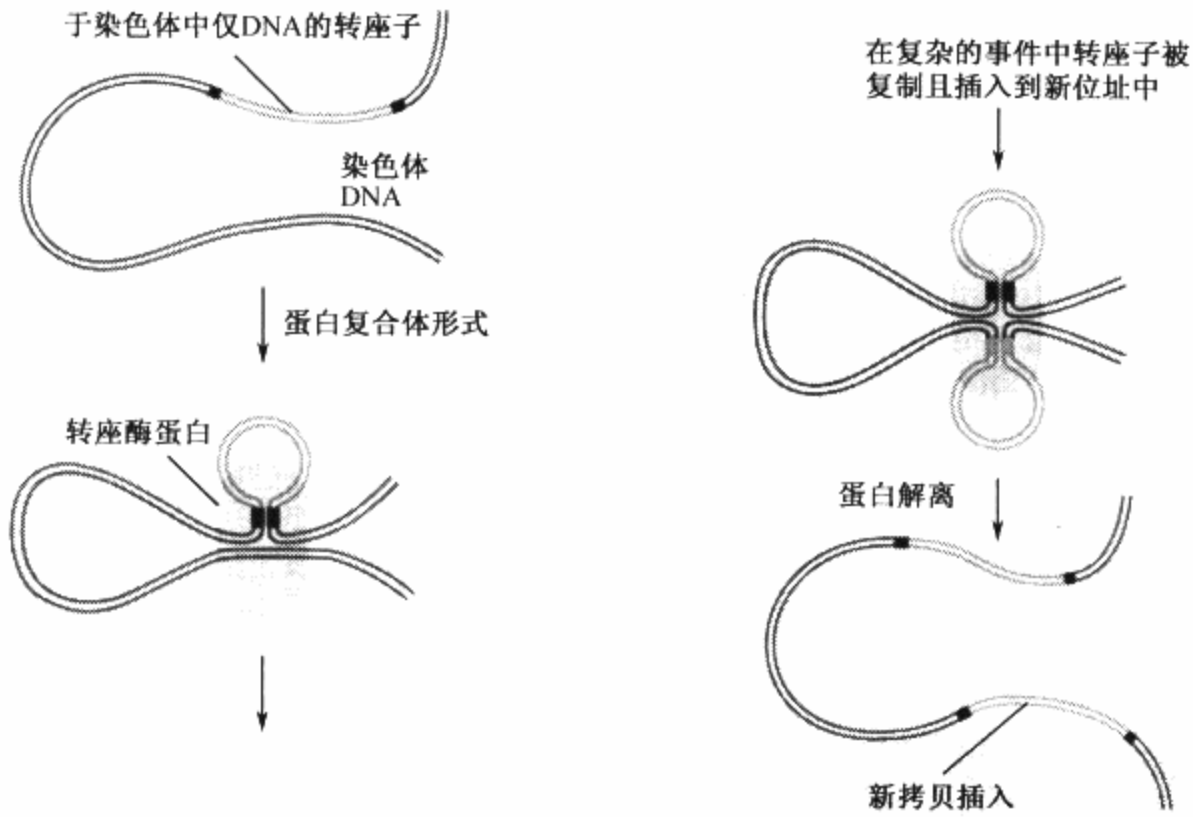


图 5-72　复制转座

在复制转座过程中，通过 DNA 复制拷贝转座子的 DNA 序列。终产物是一个与原先供体相同的 DNA 和有转座子插入其中的目标 DNA 分子。一般而言，特定的 DNA 转座子或者通过剪切-粘贴途径转座，或者通过这里概括的复制途径转座。然而，这两个途径具有许多酶的相似性，少数转座子能以任一种途径移动

一个短的重复。对于不同的转座子，这些侧翼直接重复序列的长度不同，所以可以很方便地作为以前发生过转移的位点特异性重组事件的标志。

当一个剪切-粘贴的 DNA 转座子从供体染色体上切除以后，就在空着的染色体上形成一

个双链断裂。这个断裂可以由同源末端连接反应完美地“愈合”。同样，这个断裂也可由非同源末端连接反应重新封闭起来。在这种情况下，会改变转座子的侧翼 DNA 序列，在染色体上 DNA 切除的位点上产生突变（图 5-53）。这些 DNA 转座子利用剪切-粘贴机制的变异进行了转移，称为复制转座。在这种情况下，转座子 DNA 进行复制，然后一份拷贝插入一个新的染色体位点，而最初的染色体保持不变（图 5-72）。虽然所使用的机制更加复杂，但是它与刚讨论的剪切-粘贴机制密切相关；实际上，有些转座子可以其中任何一种方式转移。

有些病毒利用转座的位点特异性重组将它们自身转移到宿主细胞染色体中

某些病毒被认为是高度可移动遗传元件，因为它们利用转座机制将它们的基因组整合到宿主细胞中。然而，这些病毒也编码蛋白质，将它们的遗传信息包装成能够感染其他细胞的病毒颗粒。许多能将自身插入宿主染色体中的病毒依靠表 5-3 所列的前两种机制来实现这一目的。实际上，我们关于这些机制的大部分了解都来源于对使用这些机制的特定病毒的研究。

感染细菌的病毒称为**噬菌体**（bacteriophage）。噬菌体 Mu 不仅利用基于 DNA 的转座将自身基因组整合到宿主染色体中，也可以利用转座过程起始它的病毒 DNA 复制。Mu 转座酶是第一个以活性形式得以钝化的转座酶，并分析了特定特征；它通过特异性地结合病毒 DNA，识别该 DNA 每个末端的重组位点，因而与刚才描述的转座酶很相似。

转座也在许多其他病毒的生命周期中发挥重要作用。其中最有名的就是**逆转录病毒**（retrovirus），包括感染人类细胞的 AIDS 病毒，称为 HIV。在细胞外，逆转录病毒以单链 RNA 基因组的形式存在，与病毒编码的**逆转录酶**（reverse transcriptase）一起，被包装到蛋白质衣壳中。在感染过程中，病毒 RNA 侵入细胞然后在逆转录酶的作用下转变成双链 DNA 分子，逆转录酶能以 RNA 或者 DNA 为模板催化 DNA 的聚合（图 5-73 和图 5-74）。逆转录病毒这个术语是因为这些病毒能逆转通常的遗传信息的流向，即从 DNA 到 RNA（图 1-5）。

然后逆转录酶产生的双链 DNA 的 2 个末端附近的特殊 DNA 序列由一种病毒编码的整合酶维系在一起。这种整合酶产生活化的 3′-OH 病毒 DNA 末端，能直接攻击目标 DNA 分子，其机制与剪切-粘贴 DNA 转座子使用的机制非常相似（图 5-75）。事实上，对细菌转座酶和 HIV 整合酶的三维结构进行详细研究后表明，这些酶存在显著的相似性，虽然人们的氨基酸序列差别相当大。

逆转录病毒样的逆转座子与逆转录病毒类似，但是缺少蛋白质衣壳

逆转录病毒将自身移入或移出染色体所使用的机制与称作**逆转录病毒样的逆转座子**（retroviral-like retrotransposon）的一个转座子大家族使用的机制相同（表 5-3）。这些元件存在于酵母、果蝇和哺乳动物等生物。其中一个研究最清楚的是酵母中发现的 Ty1 元件。与逆转录病毒一样，这种元件转座的第一步是对整个转座子进行转录，产生这种长达 5000 多核苷酸的元件的一份 RNA 拷贝。这个转录产物编码了一个逆转录酶，可由宿主细胞作为信使 RNA 进行翻译。所编码的这种酶可经过一个 RNA/DNA 杂和中间物分子，合成 RNA 分子的一个双链 DNA 拷贝，与逆转录病毒感染早期的过程几乎一模一样（图 5-73）。与逆转录病毒一样，线性双链 DNA 分子接着使用 Ty1 DNA 编码的整合酶整合到染色体上的一个位点（图 5-75）。虽然与逆转录病毒相似程度很显著，但是与它不一样的是，Ty1 元件缺少具有功能的蛋白质衣壳，因此它只能在单个细胞及其后代细胞中转移。

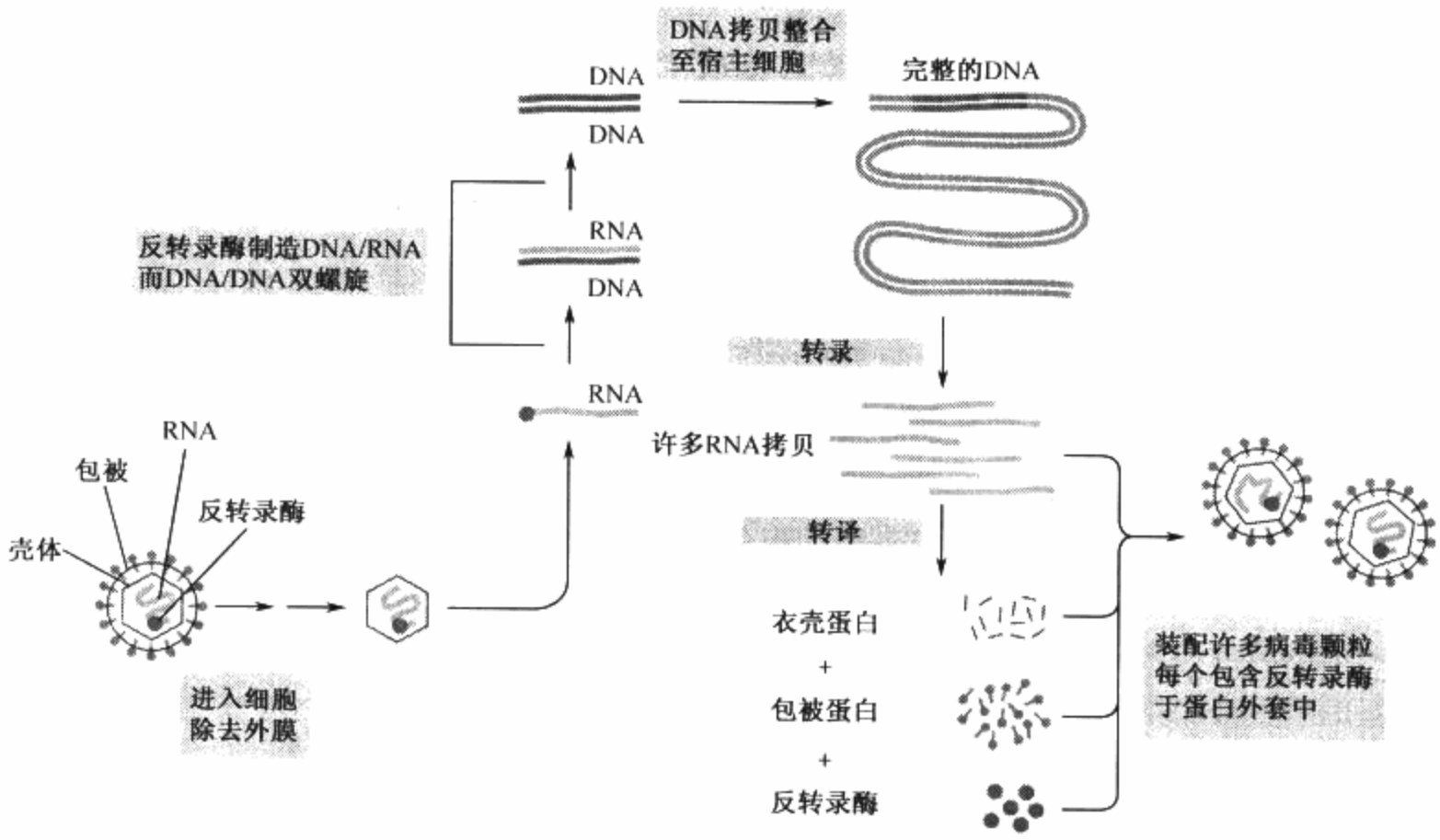


图 5-73　反转录病毒的生命周期

反转录病毒基因组由一条大约 8500 核苷酸长的 RNA 分子组成；每个病毒颗粒中包装两个这样的分子。反转录酶首先合成病毒 RNA 分子的 DNA 拷贝，然后合成第二条 DNA 链，从而产生 RNA 基因组的双链 DNA 拷贝。然后由病毒编码的整合酶催化整合到宿主染色体中（图 5-75）。这个整合过程是合成新的病毒 RNA 分子所必需的，此过程由宿主细胞 RNA 聚合酶完成，它是一种将 DNA 转录成 RNA 的酶（讨论见第 6 章）

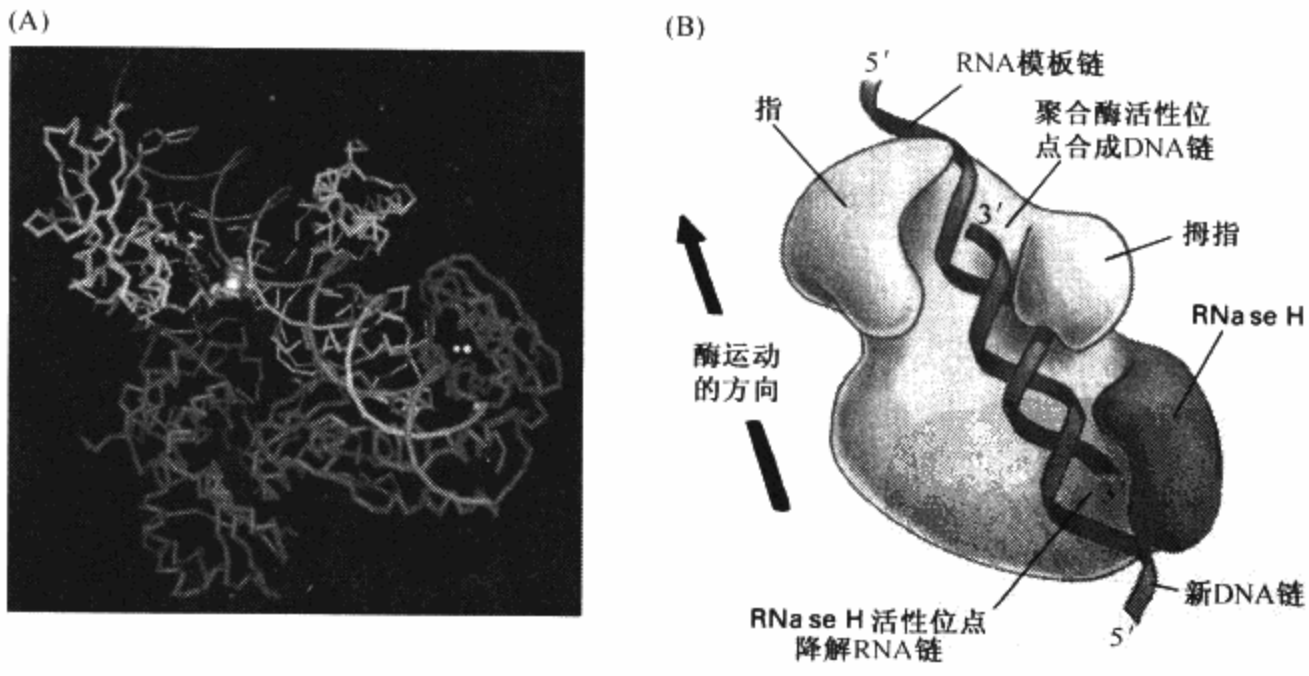


图 5-74　逆转录酶

（A）利用 X 射线晶体学得到的来自 HIV（人类艾滋病病毒）的逆转录酶的三维结构。（B）显示酶作用于 RNA 模板的模型。注意聚合酶区域（黄色）具有共价结合的 RNase H（H 指杂种）结构域（红色），它降解 RNA/DNA 螺旋中的 RNA 链。这个活性帮助聚合酶将原先的杂种螺旋转变为一个 DNA 双螺旋［(A)经 Tom Steitz 允许后使用；(B) 摘自 L. A. Kohlstaedt et al.,*Science* 256：1783～1790，1990］

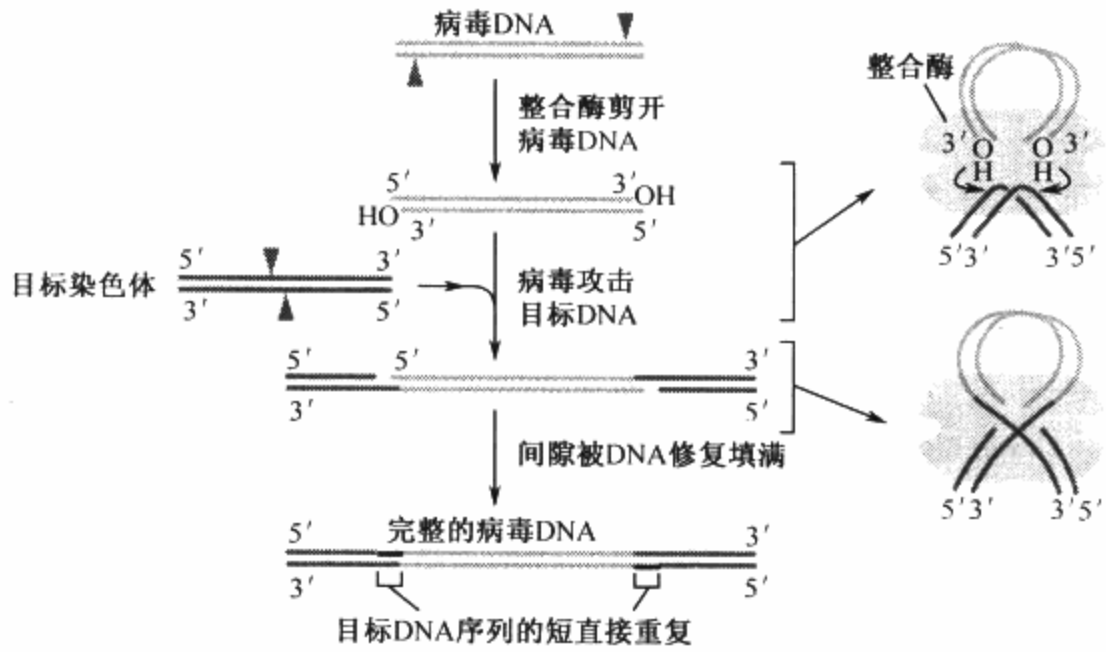


图 5-75　反转录病毒或反转录病毒样的逆转座子进行的转座的特异性重组

导致反转录病毒（如 HIV）或反转录病毒样的逆转座子（如 Ty1）的线性双链 DNA（橘黄色）整合到宿主细胞染色体（蓝色）的链断裂和链再连接过程的概要。在起始步骤中，整合酶形成一个 DNA 环并在病毒 DNA 每个末端切开一条链，暴露出一个突出的 3′-OH 基团。然后这些 3′-OH 端中的每一个直接攻击目的染色体（红色箭头）的随机选择位点的相反链上的磷酸二酯键。这将病毒 DNA 序列插入到目的染色体，在每一边留下短的缺口，这些缺口由 DNA 修复过程填补。由于存在缺口填补过程，所以这类机制（就像剪切和粘贴的转座子那样）在整合 DNA 片段的两侧留下短的目的 DNA 序列重复（黑色）；根据整合酶的不同，它们长为 3～12 个核苷酸

人类基因组大部分是由非反转录病毒的逆转座子组成

许多脊椎动物染色体很大一部分都是重复的 DNA 序列。在人类染色体中，这些重复大部分是一种称为 L1 元件（有时也称为 LINE 或长分散核元件）的逆转座子发生突变并截断后的版本。尽管大部分 L1 元件拷贝是不移动的，但仍然有少量保留了移动的能力。已经鉴定到了该元件的转座，其中一些引起人类疾病。例如，因 L1 插入编码一种血凝因子，因子Ⅷ的基因内而产生的一种特殊类型的血友病。相关的移动元件也在其他哺乳动物和昆虫，以及酵母线粒体中找到。这些非反转录病毒的逆转座子（表 5-3 中第三类）通过一种不同的机制移动，该机制需要一个由一种内切核酸酶和一种反转录酶组成的复合体。如图 5-76 所示，RNA 和反转录酶在重组中发挥的直接作用要远远大于在上面所描述的移动元件转移中发挥的作用。

在其自身核苷酸序列中不能编码内切核酸酶或逆转录酶的其他重复 DNA，被认为能够通过类似的机制，使用细胞内存在的各种内切核酸酶和反转录酶，包括那些由 L1 元件编码的酶，在染色体中增殖。例如，广泛存在的 ALu 元件缺乏内切核酸酶反转录酶，但是它仍然能增殖，并成为人类基因组的主要成分（图 5-77）。

L1 和 Alu 元件似乎在人类染色体中相当晚才增殖，因此，例如，小鼠含有与 L1 和 Alu 密切相关的序列，但是它所在小鼠染色体中的布置与在人类染色体中的布置大不一样（图 5-78）。

在不同生物中存在不同的占优势地位的转座元件

我们已经描述了几种类型的转座元件：①DNA 转座子，它的移动只涉及 DNA 断裂和连

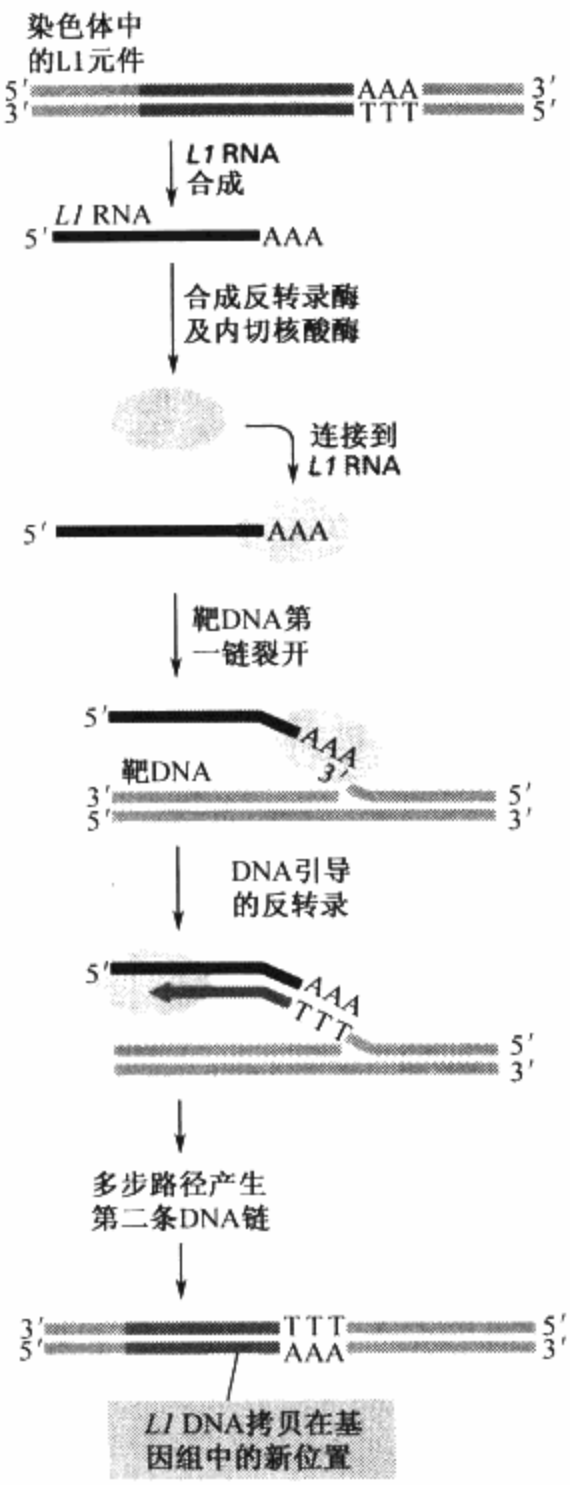


图 5-76　非反转录病毒的逆转座子进行的转座的特异性重组

L1 元件（红色）转座开始于一个内切核酸酶附着到 L1 反转录酶和 L1 RNA（蓝色），在插入将要发生的位点目标 DNA 中产生一个切口。整个切割释放出目标 DNA 的 3′-OH DNA 末端，然后它用作图中所示逆转录步骤中的引物。这产生直接与目标 DNA 相连的该元件的一个单链 DNA 拷贝。在随后的还未详细了解清楚反应中，单链 DNA 拷贝进一步加工导致产生一个插入到最初形成的切口位点的 L1 元件的新的双链 DNA 拷贝

接；②逆转录病毒样的逆转座子，它也经由 DNA 断裂和连接进行转移，但是其间 RNA 作为模板，在生成 DNA 重组底物中发挥重要作用；③非逆转录病毒的逆转座子，其中该元件的 RNA 拷贝对于它整合到目标 DNA 中起中心作用，该 RNA 作为直接的模板，参与目标 DNA 引导的逆转录过程。

有趣的是，似乎不同的生物中存在不同类型的占优势地位的转座子。例如，细菌转座子大部分都是 DNA 转座子，还存在少量的与非逆转录病毒的逆转座子相关的转座子。在酵母中，观察到的主要的可移动元件是逆转录病毒样的逆转座子。在果蝇中，这三类转座子都能找到。最后，人类基因组含有全部三种类型转座子，但是正如下面讨论的，它们的进化历史却显著不同。

基因组序列揭示出转座元件发生过转移的大概时间

人类基因组的核苷酸序列提供了跨越进化历史后转座子活动的丰富的“化石纪录”。仔细比较人类基因组中将近 3 百万多种的转座元件残迹的核苷酸序列后，可能大致重建转座子经历几十亿年的时间，在我们祖先基因组中的移动情况。例如，DNA 转座子在人类和古代猴子（25 百万～35 百万年前）开始趋异进化之前，似乎曾非常活跃；但是由于它们逐渐累积了无活性的突变，所以自那时开始，它们在人类谱系中就已失活了。同样，虽然我们基因组存在许多反转录病毒样的逆反转座子的残迹，但是现在它们似乎都已没有活性。人们认为自从约 7 百万年前，人类和大猩猩趋异进化以来，只有一个家族的反转录病毒的逆转座子在人类基因组中发生了转座。非反转录病毒的逆转座子同样非常古老，但与其他两种类型相比，它们中有些仍然在我们基因组中转移。前面提到过，它们导致了人类基因一部分新的突变，每 1000 个突变中约有 2 个。

小鼠中的情况则很不一样。虽然人类和小鼠基因组含有密度大致相同的三种类型转座子，但是两种逆转座子在小鼠基因组中仍然活跃地发生转座，导致约 10%的新突变。显然，我们还只是刚开始了解转座子的移运是如何形成现代哺乳动物的基因组。有人提出约 170 百万年前，哺乳动物自同一祖先繁衍，扩展成一个谱系的过程中，转座活动的突然暴发参与了关键的物种形成事件。在这一点上，我们只能想知道我们人类独特的品质中的许多品质是如何由许多可移动遗传元件的过去活动所产生的，现在在我们染色体中能找到这些可移动遗传元件的残迹。

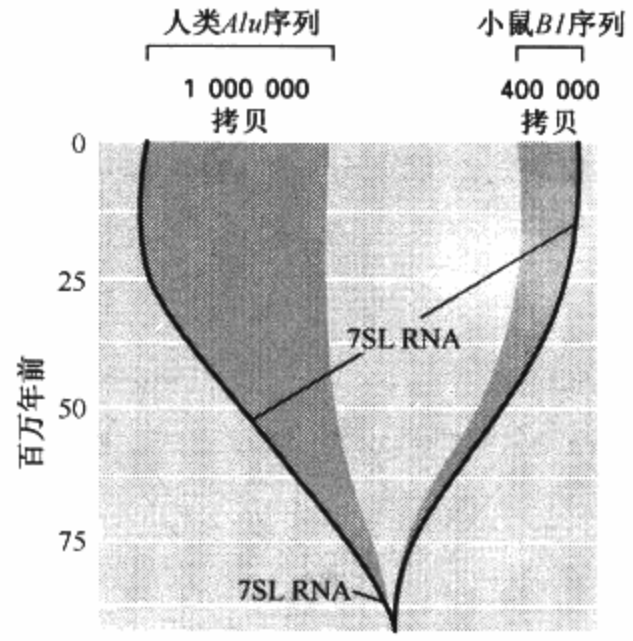


图 5-77　被建议的分别在人类和小鼠基因组中发现的富余 *Alu* 和 *B1* 序列的扩增模式

这两个转座 DNA 序列被认为从编码 SRP RNA 的重要的 7SL RNA 基因进化而来（图 12-41）。在这些高度重复元件的物种分布和序列相似性基础上，拷贝数主要扩增在小鼠和人类中似乎单独进行（图 5-78）（摘自 P. Ldeininger and E. R. Daniels, *Trends Genet*. 2：76～80，1986 和国际人类基因组测序共联体，*Nature* 409：860～921，2001）

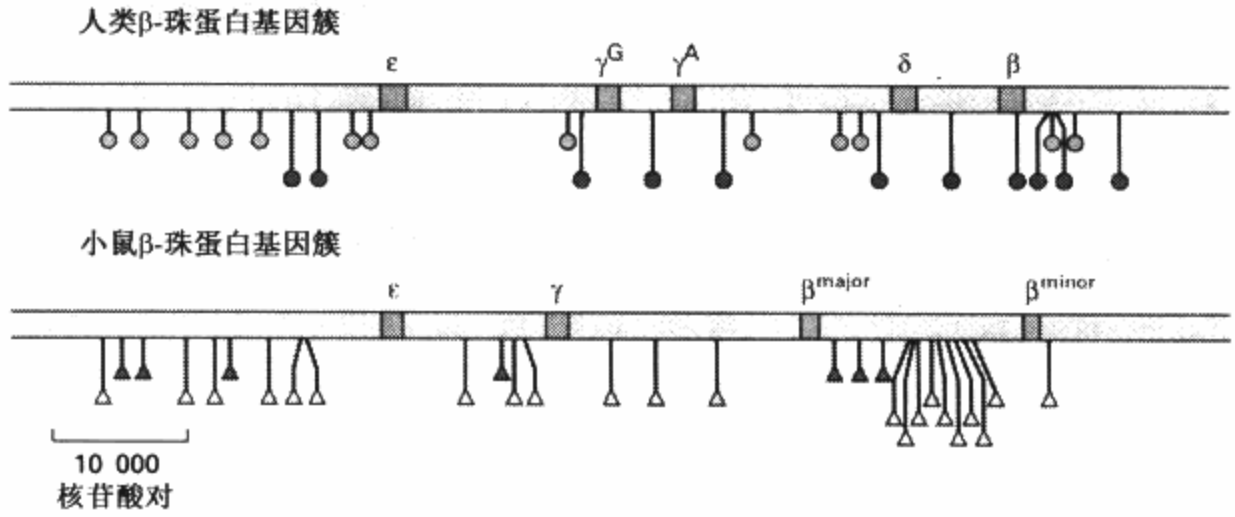


图 5-78　人类和小鼠基因组中 *β*-珠蛋白基因簇的比较，显示转座元件的位置

这幅简图中人类基因组包含 5 个有功能的 *β*-珠蛋白样基因（橙色）；而小鼠基因组中可比较的部分只有 4 个。绿色环表示人类 Alu 序列的位置，红色环表示人类 L1 序列的位置。小鼠基因组含有不同但相关的转座元件：蓝色三角表示与人类 Alu 相关的 B1 元件的位置。黄色三角表示小鼠 L1 元件的位置（与人类 L1 序列相关）。由于在小鼠和人类 β-珠蛋白基因簇中发现的转座子元件的 DNA 序列和位置如此不同，所以相信它们在相对较晚的进化时期，在各自的基因组中独立积累（经 Ross Hardison 和 Webb Miller 允许后使用）

保守的位点特异性重组能可逆地重排 DNA

一种不同类型的称为保守的位点特异性重组的特异性重组类型介导了其他类型的可移动 DNA 元件的重排。在这种途径中，断裂和连接发生在两个特殊位点，每条参与的 DNA 分子上各有一个。根据两个重组位点的方向，能发生 DNA 整合、DNA 切除或者 DNA 倒位（图 5-79）。

位点特异性重组酶在每条 DNA 分子的特殊序列上断裂并重新连接两个 DNA 双螺旋，这

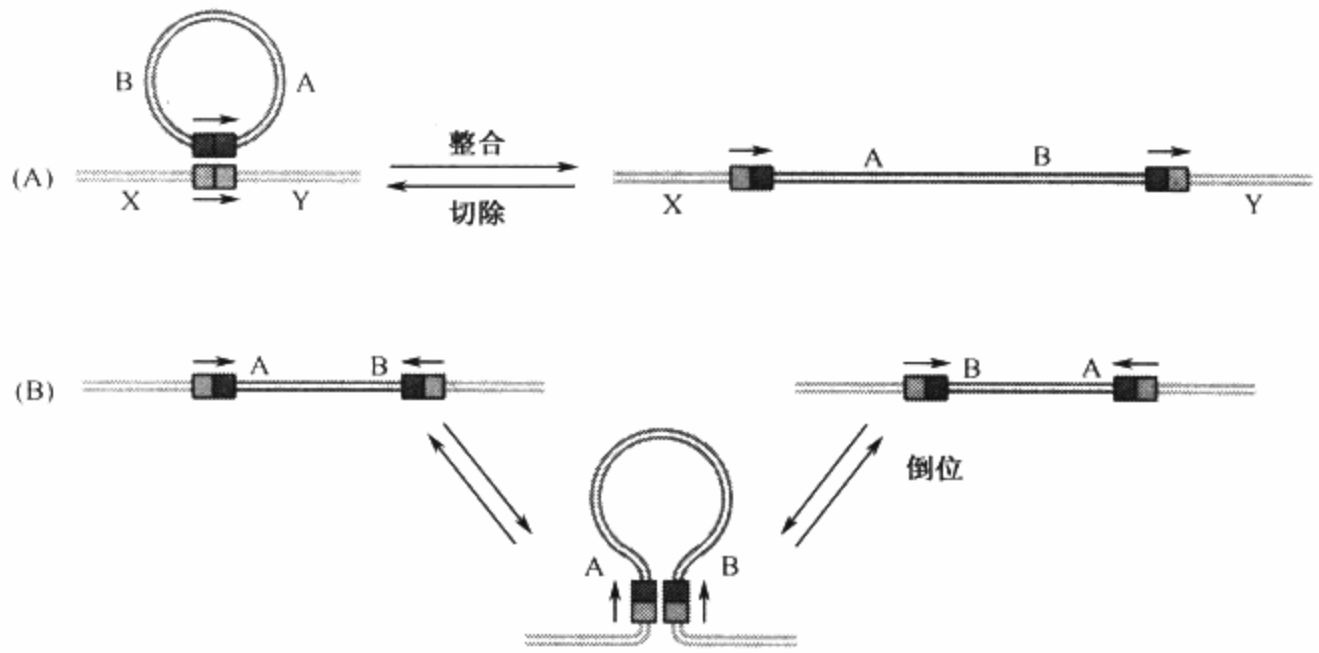


图 5-79　保守的位点特异重组产生的两种 DNA 重排

（A）和（B）中的反应之间唯一差别在于发生位点特异性重组的两个 DNA 位点（用箭头表示）相对方向不同。（A）通过整合反应，一个环状 DNA 分子能整合到另一个 DNA 分子中；通过逆反应（切除），它能出来重新形成原先的 DNA 环。λ噬菌体和其他细菌病毒正是以这种方式从它们的宿主染色体进进出出的。（B）保守的位点特异性重组也能倒位染色体中的特定 DNA 片段。一个详细研究的 DNA 倒位的例子，发生在鼠伤寒沙门杆菌中，它是人类中引起食物中毒主要原因的一种生物，DNA 片段的倒位改变了细菌产生的鞭毛类型（图 7-64）

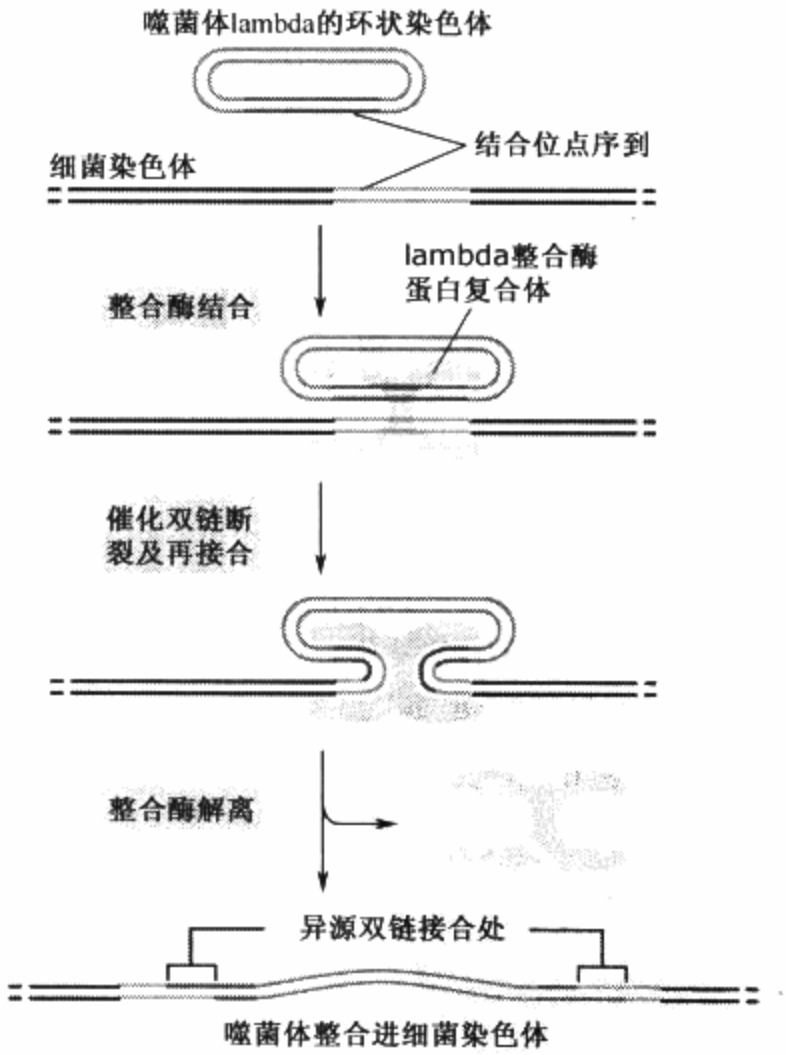


图 5-80　环状 λ 噬菌体 DNA 染色体插入细菌染色体中

在这个位点特异重组中，λ整合酶结合每个染色体上特殊的“附着位点”DNA 序列，并在此切开并将小段同源 DNA 序列包括在内。然后整合酶将伙伴链打开，再重新封闭以形成一个 10 个核苷酸对长的异源双链连接，总共需要 4 个链断裂和链连接反应；对于它们每一个而言，断裂的磷酸二酯键的能量储存在 DNA 和酶之间短暂的共价连接中，这样 DNA 链的重新封闭就不需 ATP 或 DNA 连接酶

些酶常常以可逆的方式实现这一点：将两条 DNA 分子连接起来的同一个酶系统能再次使它们分离并精确地恢复成原来的两条 DNA 分子。因此，将这种类型的重组称为“保守”的位点特异性重组，以便与前面讨论的机制不同的转座的位点特异性重组区别。

一种细菌病毒，λ 噬菌体，是从生物化学方面第一个详细研究的可移动 DNA 元件。当该病毒进入细胞时，就合成一种病毒编码的酶，该酶称为 λ 整合酶。该酶介导病毒 DNA 与细菌染色体的共价连接，使病毒成为细菌染色体的一部分，这样它就能自动地进行复制，作为宿主 DNA 的一部分。λ 整合酶反应的一个重要特征就是重组位点由两条相关但不同的 DNA 序列的识别决定，一条定位于噬菌体染色体，而另一条位于宿主细菌的染色体。当几个整合酶蛋白质分子以及几个宿主蛋白质牢固地结合到噬菌体环状染色体的一段特殊 DNA 序列上的时候，重组过程就开始了。该 DNA-蛋白质复合物现在就能结合到细菌染色体上的一段附着位点 DNA 序列，从而将细菌和噬菌体染色体带到一起。然后，整合酶催化所需的切割和填补反应，引起位点特异的链交换。因为两条连接序列存在一小段同源序列区域，所以在这个交换位点形成一段小的异源双链连接（图 5-80）。

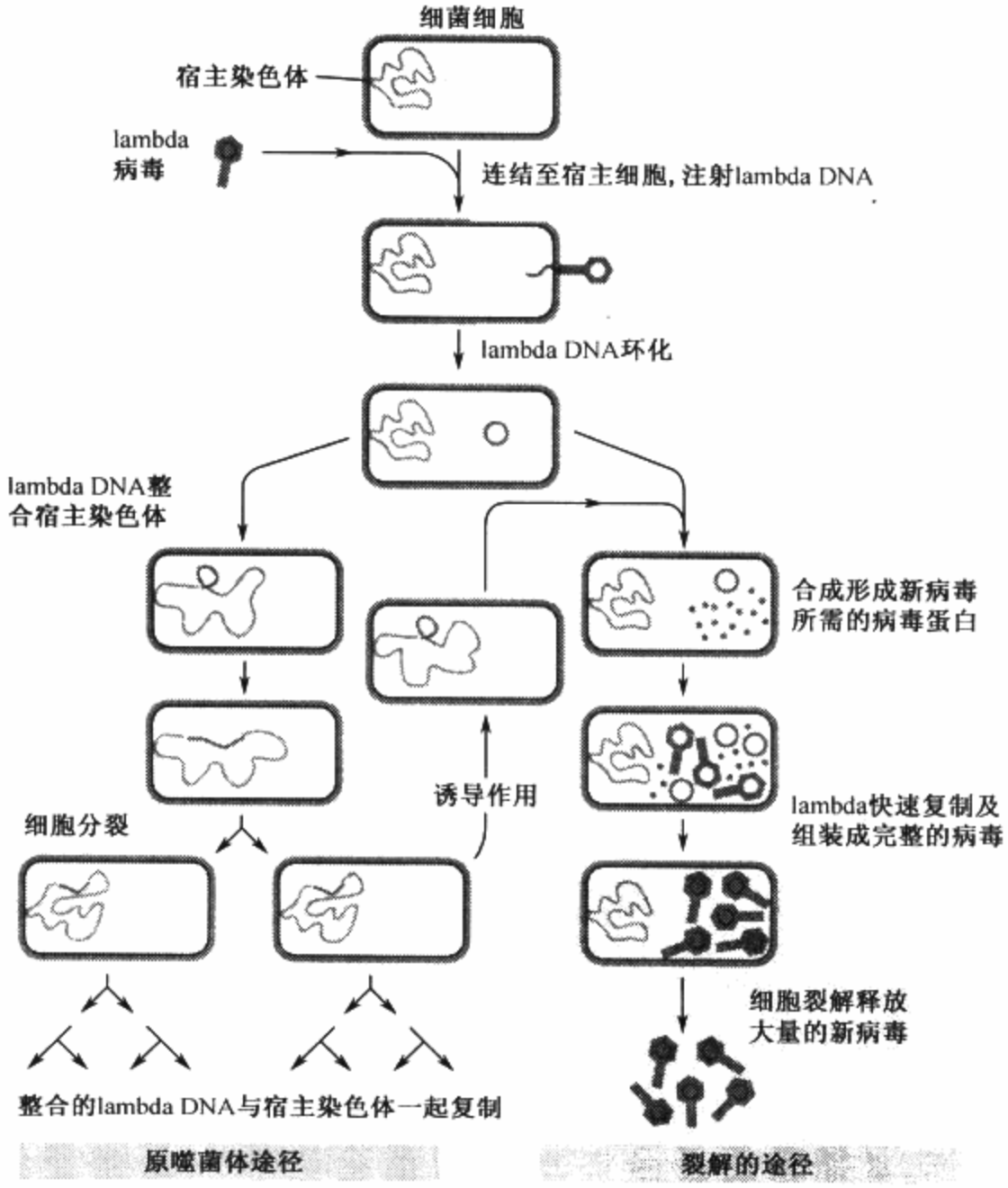


图 5-81　λ 噬菌体的生命周期

双链 λDNA 基因组含有 50 000 个核苷酸对，并编码 50～60 个不同的蛋白质。当 λDNA 进入细胞后，末端就连接起来形成个环状 DNA 分子。该噬菌体能在大肠杆菌中通过裂解途径增殖，它会破坏细胞，或者它能进入一种潜伏的溶原状态。携带 λ 溶原菌的细胞受到损伤，会诱导溶原菌从宿主染色体出来，并转变为裂解生长（绿色箭头）。λDNA 不论进入细菌染色体还是出来，都通过保守的位点特异性重组，由 λ 整合酶催化完成（图 5-80）

λ整合酶与DNA拓扑异构酶类似，当它将链打断时，与DNA形成一个可逆的共价连接。因此，该位点特异性重组事件即便在缺少ATP和DNA连接酶的情况下也能发生，后两者对于磷酸二酯键的形成通常是必需的。

相同类型的位点特异性重组机制也可反过来用于促进结合在特殊的重组位点上的可移动元件的切除，这些位点以正向重复的形式存在。在λ噬菌体中，在对特异信号的应答中切除使其能够从大肠杆菌染色体中的整合位点出来，并在细菌细胞内迅速增殖（图5-81）。切除由一个蛋白质复合体催化，该蛋白质复合体是由整合酶和宿主因子以及另一种噬菌体蛋白质——切除酶组成。切除酶只在宿主细胞处于胁迫时才由病毒产生，在这种情况下，噬菌体倾向于离开宿主细胞并再次作为病毒颗粒增殖。

保守的位点特异性重组能够用来打开或关闭基因

如果保守的位点特异性重组识别的特异位点方向发生倒位，那么位点间的DNA序列也会倒位，而不会切除（图5-79）。许多细菌用DNA序列的这种倒位来控制特定基因的表达。例如，通过将分开的编码片段排列成活跃的基因。这种基因控制具有可直接遗传的优势，因为当细胞分裂时，DNA的新排列自动转移到子代染色体中。

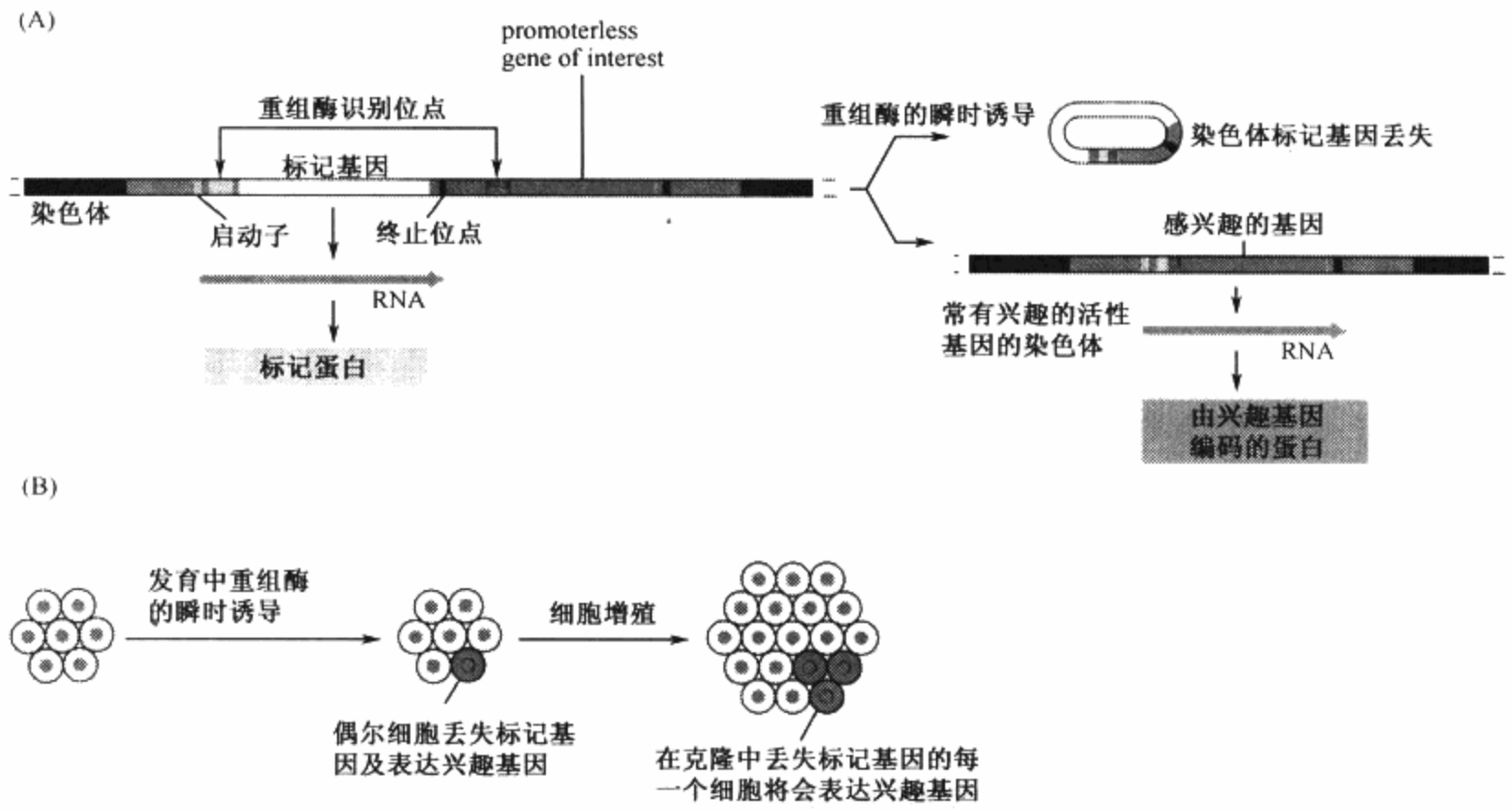


图5-82　转基因动物中如何利用保守的位点特异性重组酶打开一群细胞特异基因

该技术需要两个专门通过基因工程方法得到的DNA分子插入到动物的生殖细胞系中。（A）图中所示的DNA分子已经过改造，具有特异的识别位点（绿色），这样感兴趣的基因（红色）只有在利用这些位点的位点特异性重组酶受诱导后才转录。如右图所示，这个诱导移走一个标记基因（黄色）并带来邻近兴趣基因的启动子DNA（橙色）。重组酶是可诱导的，因为它由另一个DNA分子（外显子）编码，该DNA已经过改造，已确保该酶只在动物用特殊的小分子处理或者升高温度之后才合成。（B）重组酶的瞬时诱导引起该酶短暂的大量合成，从而引起偶然细胞中发生DNA重排。对于该细胞以及它所有的后代而言，标记基因是失活的，兴趣基因则自发活化[如(A)中所示]。在表达兴趣基因的发育动物中细胞的克隆能通过它们标记蛋白质的丢失加以鉴定。这种技术广泛用于小鼠和果蝇，因为它使研究者能够研究在完好的动物中一组细胞表达任何兴趣基因的影响。在利用该技术的一个实例中，利用了P1噬菌体的Cre重组酶和它的loxP识别位点

这些类型的酶也成为细胞及发育学家强有力的工具。为了阐明特定基因和蛋白质在复杂的多细胞生物内的作用，可以利用遗传工程技术将编码位点特异性重组酶的基因和含有该酶识别 DNA 位点的精心设计的目标 DNA 引入小鼠中。在合适的时间，该酶的编码基因能被激活，并重新排列目标 DNA 序列。这种重排常用来在小鼠特定组织中诱导特定蛋白质的合成（图 5-82）。通过同样的方式，该技术也可用于在感兴趣的组织中关闭任何的特定基因。通过这种办法，理论上研究者可以确定出任何蛋白质对完好动物的任何组织的影响。

小结

几乎所有生物体的基因组都含有可移动遗传元件，能以转座的或保守的位点特异性重组过程从基因组中一个位点转移到另一个位点。在大多数情况下，这种转移是随机的并且发生的频率很低。可移动遗传元件包括转座子［它只能在单个细胞内移动（以及该细胞的子细胞）］和那些基因组能整合到宿主细胞基因组中去的病毒。

存在三类转座子：DNA 转座子、反转录病毒样的逆转座子以及非反转录病毒的逆转座子。除第三类外，其他两类在病毒之间都存在同源物。尽管可以将病毒和转座元件视为寄生物，但是在它们的位点特异性重组事件导致的 DNA 序列新的排列中，有许多形成了对细胞和生物体进化很关键的遗传变化。

参考文献

一般文献

Friedberg EC, Walker GC & Siede W (1995) DNA Repair and Mutagenesis. Washington, DC: ASM Press.

Hartwell L, Hood L, Goldberg ML et al. (2000) Genetics: from Genes to Genomes. Boston: McGraw Hill.

Kornberg A & Baker TA (1992) DNA Replication, 2nd edn. New York: WH Freeman.

Lodish H, Berk A, Zipursky SL et al. (2000) Molecular Cell Biology, 4th edn. New York: WH Freeman.

Stent GS (1971) Molecular Genetics: An Introductory Narrative. San Francisco: WH Freeman.

The Maintenance of DNA Sequences

Crow JF (2000) The origins, patterns and implications of human spontaneous mutation. *Nat. Rev. Genet.* 1, 40–47.

Drake JW, Charlesworth B, Charlesworth D & Crow JF (1998) Rates of spontaneous mutation. *Genetics* 148, 1667–1686.

Ohta T & Kimura M (1971) Functional organization of genetic material as a product of molecular evolution. *Nature* 233, 118–119.

Wilson AC & Ochman H (1987) Molecular time scale for evolution. *Trends Genet.* 3, 241–247.

DNA Replication Mechanisms

Arezi B & Kuchta RD (2000) Eukaryotic DNA primase. *Trends Biochem. Sci.* 25, 572–576.

Baker TA & Bell SP (1998) Polymerases and the replisome: machines within machines. *Cell* 92, 295–305.

Buermeyer AB, Deschenes SM, Baker SM & Liskay RM (1999) Mammalian DNA mismatch repair. *Annu. Rev. Genet.* 33, 533–564.

Davey MJ & O'Donnell M (2000) Mechanisms of DNA replication. *Curr. Opin. Chem. Biol.* 4, 581–586.

Kelman Z & O'Donnell M (1995) DNA polymerase III holoenzyme: structure and function of a chromosomal replicating machine. *Annu. Rev. Biochem.* 64, 171–200.

Kolodner RD (2000) Guarding against mutation. *Nature* 407, 687, 689.

Kornberg A (1960) Biological synthesis of DNA. *Science* 131, 1503–1508.

Kunkel TA & Bebenek K (2000) DNA replication fidelity. *Annu. Rev. Biochem.* 69, 497–529.

Kuriyan J & O'Donnell M (1993) Sliding clamps of DNA polymerases. *J. Mol. Biol.* 234, 915–925.

Li JJ & Kelly TJ (1984) SV40 DNA replication *in vitro*. *Proc. Natl. Acad. Sci. USA* 81, 6973.

Marians KJ (2000) Crawling and wiggling on DNA: structural insights to the mechanism of DNA unwinding by helicases. *Structure Fold. Des.* 8, R227–R235.

Meselson M & Stahl FW (1958) The replication of DNA in *E. coli*. *Proc. Natl Acad. Sci. USA* 44, 671–682.

Modrich P & Lahue R (1996) Mismatch repair in replication fidelity, genetic recombination, and cancer biology. *Annu. Rev. Biochem.* 65, 101–133.

Ogawa T & Okazaki T (1980) Discontinuous DNA replication. *Annu. Rev. Biochem.* 49, 421–457.

Okazaki R, Okazaki T, Sakabe K et al. (1968) Mechanism of DNA chain growth. I. Possible discontinuity and unusual secondary structure of newly synthesized chains. *Proc. Natl. Acad. Sci. USA* 59, 598–605.

Postow L, Peter BJ & Cozzarelli NR (1999) Knot what we thought before: the twisted story of replication. *Bioessays* 21, 805–808.

von Hippel PH & Delagoutte E (2001) A general model for nucleic acid helicases and their "coupling" within macromolecular machines. *Cell* 104, 177–190.

Waga S & Stillman B (1998) The DNA replication fork in eukaryotic cells. *Annu. Rev. Biochem.* 67, 721–751.

Wang JC (1996) DNA topoisomerases. *Annu. Rev. Biochem.* 65, 635–692.

Young MC, Reddy MK & von Hippel PH (1992) Structure and function of the bacteriophage T4 DNA polymerase holoenzyme. *Biochemistry* 31, 8675–8690.

The Initiation and Completion of DNA Replication in Chromosomes

DePamphilis ML (1999) Replication origins in metazoan chromosomes: fact or fiction? *Bioessays* 21, 5–16.

Donaldson AD & Blow JJ (1999) The regulation of replication origin activation. *Curr. Opin. Genet. Dev.* 9, 62–68.

Dutta A & Bell SP (1997) Initiation of DNA replication in eukaryotic cells. *Annu. Rev. Cell Dev. Biol.* 13, 293–332.

Echols H (1990) Nucleoprotein structures initiation DNA replication, transcription and site-specific recombination. *J. Biol. Chem.* 265, 14697–14700.

Gasser SM (2000) A sense of the end. *Science* 288, 1377–1379.

Heun P, Laroche T, Raghuraman MK & Gasser SM (2001) The positioning and dynamics of origins of replication in the budding yeast nucleus. *J. Cell Biol.* 152, 385–400.

Huberman JA & Riggs AD (1968) On the mechanism of DNA replication in mammalian chromosomes. *J. Mol. Biol.* 32, 327–341.

Kass-Eisler A & Greider CW (2000) Recombination in telomere-length maintenance. *Trends Biochem. Sci.* 25, 200–204.

McEachern MJ, Krauskopf A & Blackburn EH (2000) Telomeres and their control. *Annu. Rev. Genet.* 34, 331–358.

Newlon CS & Theis JF (1993) The structure and function of yeast ARS elements. *Curr. Opin. Genet. Dev.* 3, 752–758.

Newlon CS (1997) Putting it all together: building a prereplicative complex. *Cell* 91, 717–720.

Randall SK & Kelly TJ (1992) The fate of parental nucleosomes during SV40 DNA replication. *J. Biol. Chem.* 267, 14259–14265.

Russev G & Hancock R (1982) Assembly of new histones into nucleosomes and their distribution in replicating chromatin. *Proc. Natl Acad. Sci. USA* 79, 3143–3147.

Wintersberger E (2000) Why is there late replication? *Chromosoma* 109, 300–307.

DNA Repair

Auerbach AD & Verlander PC (1997) Disorders of DNA replication and repair. *Curr. Opin. Pediatr.* 9, 600–616.

Critchlow SE & Jackson SP (1998) DNA end-joining: from yeast to man. *Trends Biochem. Sci.* 23, 394–398.

Lindahl T (1993) Instability and decay of the primary structure of DNA. *Nature* 362, 709–715.

Lindahl T, Karran P & Wood RD (1997) DNA excision repair pathways. *Curr. Opin. Genet. Dev.* 7, 158–169.

O'Connell MJ, Walworth NC & Carr AM (2000) The G2-phase DNA-damage checkpoint. *Trends Cell Biol.* 10, 296–303.

Parikh SS, Mol CD, Hosfield DJ et al. (1999) Envisioning the molecular choreography of DNA base excision repair. *Curr. Opin. Struct. Biol.* 9, 37–47.

Perutz MF (1990) Frequency of abnormal human haemoglobins caused by C→T transitions in CpGdinucleotides. *J. Mol. Biol.* 213, 203–206.

Sutton MD, Smith BT, Godoy VG et al. (2000) The SOS response: recent insights into umuDC-dependent mutagenesis and DNA damage tolerance. *Annu. Rev. Genet.* 34, 479–497.

Walker GC (1995) SOS-regulated proteins in translesion DNA synthesis and mutagenesis. *Trends Biochem. Sci.* 20, 416–420.

General Recombination

Baumann P & West SC (1998) Role of the human RAD51 protein in homologous recombination and double-stranded-break repair. *Trends Biochem. Sci.* 23, 247–251.

Flores-Rozas H & Kolodner RD (2000) Links between replication, recombination and genome instability in eukaryotes. *Trends Biochem. Sci.* 25, 196–200.

Haber JE (1998) Mating-type gene switching in Saccharomyces cerevisiae. *Annu. Rev. Genet.* 32, 561–599.

Haber JE (2000) Recombination: a frank view of exchanges and vice versa. *Curr. Opin. Cell Biol.* 12, 286–292.

Holliday R (1990) The history of the DNA heteroduplex. *Bioessays* 12, 133–142.

Kowalczykowski SC & Eggleston AK (1994) Homologous pairing and DNA strand-exchange proteins. *Annu. Rev. Biochem.* 63, 991–1043.

Kowalczykowski SC (2000) Initiation of genetic recombination and recombination-dependent replication. *Trends Biochem. Sci.* 25, 156–165.

Szostak JW, Orr-Weaver TK, Rothstein RJ et al. (1983) The double-strand break repair model for recombination. *Cell* 33, 25–35.

van Gent DC, Hoeijmakers JH & Kanaar R (2001) Chromosomal stability and the DNA double-stranded break connection. *Nat. Rev. Genet.* 2, 196–206.

Welcsh PL, Owens KN & King MC (2000) Insights into the functions of BRCA1 and BRCA2. *Trends Genet.* 16, 69–74.

West SC (1997) Processing of recombination intermediates by the RuvABC proteins. *Annu. Rev. Genet.* 31, 213–244.

Site-specific Recombination

Campbell AM (1993) Thirty years ago in genetics: prophage insertion into bacterial chromosomes. *Genetics* 133, 433–438.

Craig NL (1996) Transposition, in *Escherichia coli* and *Salmonella*, pp 2339–2362. Washington, DC: ASM Press.

Craig NL (1997) Target site selection in transposition. *Annu. Rev. Biochem.* 66, 437–474.

Gopaul DN & Duyne GD (1999) Structure and mechanism in site-specific recombination. *Curr. Opin. Struct. Biol.* 9, 14–20.

Gottesman M (1999) Bacteriophage lambda: the untold story. *J. Mol. Biol.* 293, 177–180.

Jiang R & Gridley T (1997) Gene targeting: things go better with Cre. *Curr. Biol.* 7, R321–323.

Mizuuchi K (1992) Transpositional recombination: mechanistic insights from studies of mu and other elements. *Annu. Rev. Biochem.* 61, 1011–1051.

Sandmeyer SB (1992) Yeast retrotransposons. *Curr. Opin. Genet. Dev.* 2, 705–711.

Smith AF (1999) Interspersed repeats and other mementos of transposable elements in mammalian genomes. *Curr. Opin. Genet. Dev.* 9, 657–663.

Stark WM, Boocock MR & Sherratt DJ (1992) Catalysis by site-specific recombinases. *Trends Genet.* 8, 432–439.

Varmus H (1988) Retroviruses. *Science* 240, 1427–1435.

6　细胞如何解读基因组

从 DNA 到 RNA
从 RNA 到蛋白质
RNA 世界与生命的起源

只有在 20 世纪 50 年代早期发现了 DNA 结构以后，才知道了细胞内的遗传信息是如何在 DNA 核苷酸序列中编码的。从那以后取得的进展令人惊奇。50 年后，我们完成了许多生物，包括人类基因组的测序，因而知道了产生一个像我们自己这样复杂的生物体所需要的最大信息量。生命所需遗传信息的有限性限制了细胞的生物化学以及结构特征，也清楚地表明生物学并不是无限复杂的。

本章我们解释如何解码和利用基因组中的信息。我们将看到，人类已经了解了很多有关四字母表（DNA 中 4 种不同的核苷酸）编写的遗传指令如何指导形成细菌、果蝇，或者人。但是要弄清生物基因组中储存的信息如何形成即便是最简单的，只有 500 个基因的单细胞细菌，仍然需要我们大量的努力，更不用说这些信息如何指导形成拥有将近 30 000 个基因的人类了。目前仍然存在大量未知领域；因此，许多令人着迷的挑战正等着下一代细胞生物学家去征服。

考察一小部分果蝇基因组后，就能让我们认识到细胞在解码基因组中所面临的问题（图 6-1）。各种基因组中大部分 DNA 编码信息都用于指令生物合成的每种蛋白质的氨基酸线性顺序——序列。正如第 3 章描述的，氨基酸序列继而再指令每个蛋白质如何折叠，使分子具有不同的形状和化学性质。因此当细胞合成某种特定的蛋白质时，就必须对基因组相应的编码区域进行精确的解码。基因组 DNA 中编码的其他信息规定了每个基因表达成蛋白质，究竟在生物生命过程中哪个确切的时间以及哪个确切的细胞类型中进行。由于蛋白质是细胞主要组分，所以基因组的解码不仅决定了细胞的大小、形状、化学特性和行为，而且决定了地球上每个物种的不同特征。

可能有人会推测基因组中的信息是井井有条地排列的，就像一本字典或者一个电话簿。虽然有些细菌的基因组看起来相当有组织，但是大多数多细胞生物的基因组，如果蝇，却出人意料地杂乱无章。小段的编码 DNA（也就是编码蛋白质的 DNA）分散在大段表面上看毫无意义的 DNA 之间。基因组有些部分含有许多基因，而另一部分则完全不含基因。细胞内功能紧密相关的蛋白质的基因常常位于不同的染色体，而相邻基因编码的蛋白质通常在细胞内毫不相干。由此可见，解读基因组并不是简单的事情。即使在功能强大的电脑帮助下，研究者们仍然难以确切地定位出复杂基因组 DNA 序列中基因从哪里开始，在哪里结束，更不用说去预计每个基因在生物体生命的哪一时刻表达了。虽然已经知道了人类基因组 DNA 序列，但是可能人类仍然需要至少 10 年的时间去鉴定出每一个基因，并确定它所产生的蛋白质准确的氨基酸序列。然而我们体内的细胞每秒钟可以成千上万遍地做到这一点。

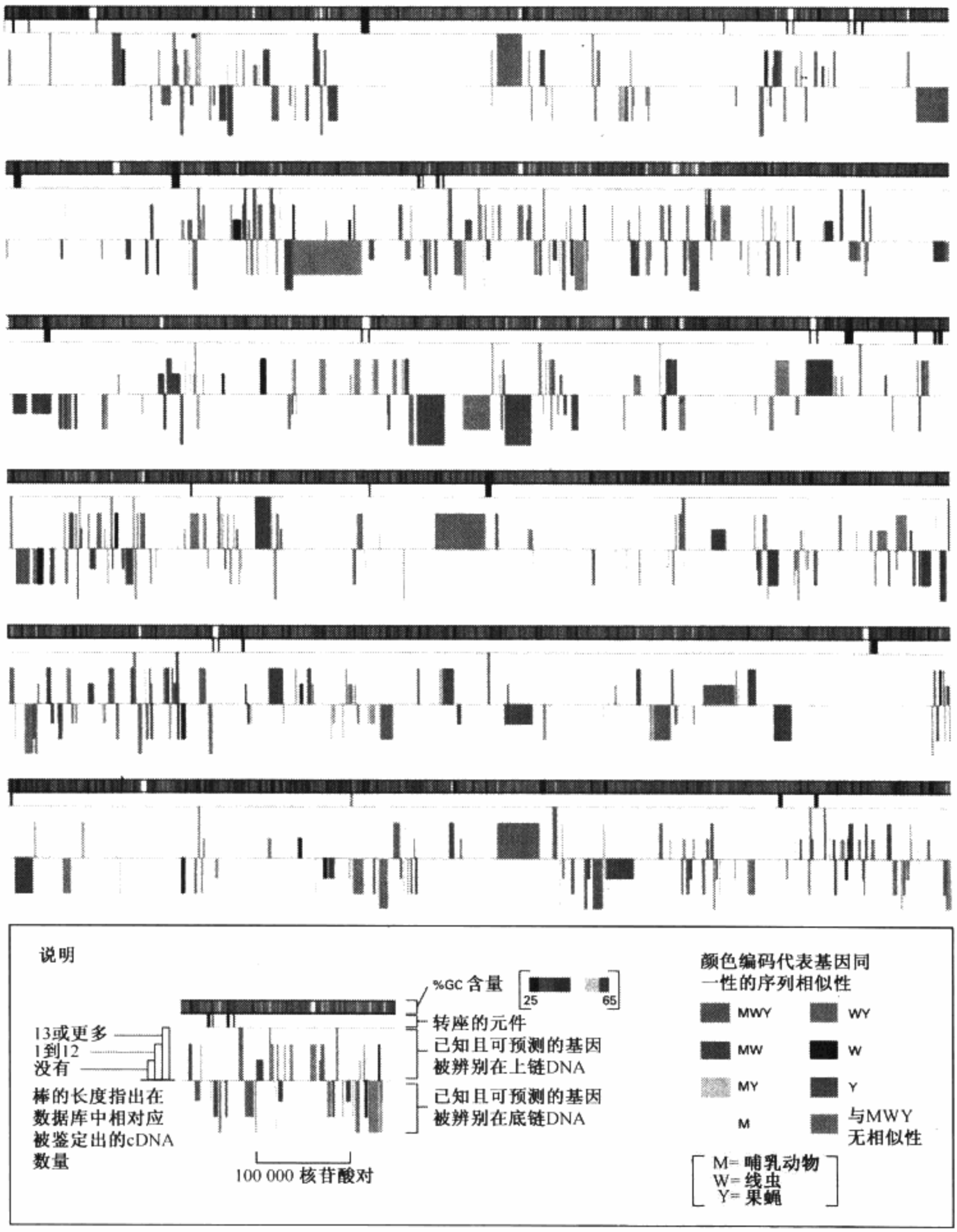


图 6-1　黑腹果蝇基因组 2 号染色体一部分的图解描述

该图代表大约整个果蝇基因组 3%，分别排列成 6 条连续的片段。正如图例中总结的，符号代表意义是：彩色的条带：G-C 碱基对成分；不同厚度的黑色垂直线：转座元件的位置，较厚的条带表示元件簇；彩色框：一条 DNA 链上编码的（位于中间线上部的框）基因（已知和预测的）和另一条链上编码的基因（中间线下面的方框）。每个预测基因的长度包括其外显子（编码蛋白的 DNA）和内含子（非编码 DNA）（图 4-25）。如图例中所示，每个基因框高度与不同数据库中匹配该基因的 cDNA 数量呈比例。第 8 章将描述，cDNA 是 mRNA 分子的 DNA 拷贝，多种数据库中已储存了大量的 cDNA 核苷酸序列的集合。cDNA 核苷酸序列与特定的预测基因核苷酸序列匹配数量越多，就可以更加确信所预测的基因转录成 RNA，因而是一个真基因。每个基因框（见图例中颜色编码）的颜色表示其他生物中是否存在一个密切相关的基因。例如，MWY 表示该基因在哺乳动物和线虫以及酿酒酵母中有近亲，MW 则表示该基因在哺乳动物和线虫中有近亲但在酵母中没有（摘自 Mark D. Adams et al., Science, 287: 2185～2195, 2000）

基因组中的DNA本身不能指导蛋白质的合成，而是利用了RNA作为中间物分子。当细胞需要特定的蛋白质时，就首先将染色体中冗长DNA分子的适当部位拷贝成RNA（此过程称为转录）。正是DNA片段的这些RNA拷贝直接用作模板，指导蛋白质的合成（此过程称为翻译）。因此遗传信息在细胞内的流向是从DNA到RNA，再到蛋白质（图6-2）。所有细胞，从细菌到人类，都以这种方式表达它们的基因，这是一个如此根本的原理，以至于称它为分子生物学中心法则。

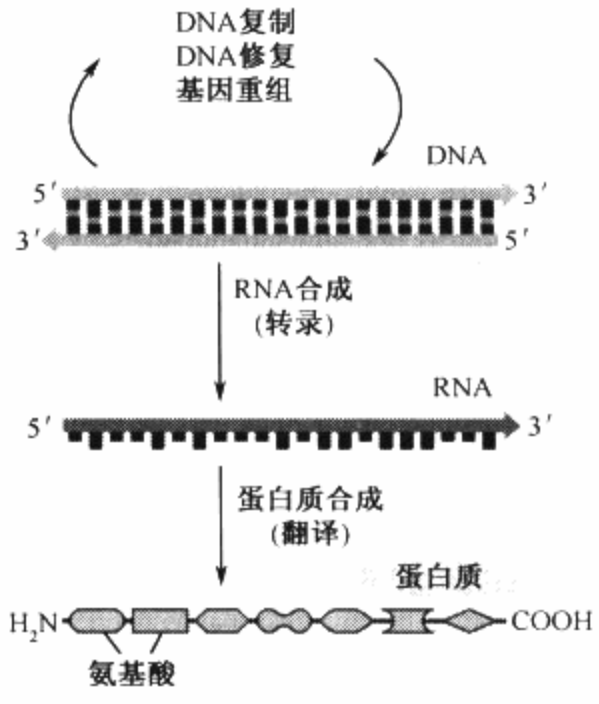


图6-2　从DNA到蛋白质的途径
在所有活细胞中都发生着遗传信息流从DNA到RNA（转录），再从RNA到蛋白质（翻译）

尽管中心法则具有普遍性，信息流从DNA到蛋白质在方式上也有许多重要的变化。其中的原理就在于，真核细胞中RNA转录物能够从细胞核内出来并翻译成蛋白质之前，在细胞核内经过一系列加工过程，包括RNA剪接。这些加工过程会关键性地改变RNA分子的“含义”。因此对于理解真核生物如何解读基因组很重要。最后，虽然我们本章主要讨论基因组编码蛋白质的合成，但是我们要看到，有些基因的终产物是RNA。与蛋白质相似，这些RNA有许多折叠成精确的三级结构，在细胞内发挥结构和催化作用。

本章我们首先讨论解码基因组的第一步：一个基因的DNA产生一个RNA分子的转录过程。然后我们按这个RNA分子在细胞中的命运继续讨论，直到最后形成正确折叠的蛋白质为止。在本章的最后，我们来了解现存的相当复杂的信息存储体系，以及转录和翻译可能是怎样从细胞进化最早期比较简单的系统衍生而来的。

从DNA到RNA

转录和翻译是细胞解读或表达它们基因中遗传指令的方式。由于从同一基因能产生许多相同的RNA拷贝，而且每个RNA分子都能指导合成许多相同的蛋白质，所以如果需要的话，细胞能迅速合成大量蛋白质。但是每个基因也能以不同的效率进行转录和翻译，使得细胞能够大量的合成某些蛋白质而又小量的合成另一些（图6-3）。此外，我们在下一章将会看到，细胞能根据当时的需要情况，改变或调节它每一个基因的表达——最明显的就是通过控制它的RNA的合成。

部分DNA序列转录成RNA

细胞解读它所需的遗传指令部分时，第一步就是将其DNA核苷酸序列特定部分（一个基因）拷贝成RNA核苷酸序列。虽然RNA中信息拷贝成另一种化学形式，但其编写语言仍然与DNA中的编写语言在本质上相同——核苷酸序列的语言。由此得到这个名字，**转录**（transcription）。

与DNA一样，RNA是由4种不同的核苷酸亚单位通过磷酸二酯键连接起来形成的线性聚合物（图6-4）。它在化学上与DNA存在两方面差别：①RNA中的核苷酸是核糖核苷酸——也就是说它们含有核糖（由此产生核糖核苷酸这个名字）而不是脱氧核糖核苷酸；②虽然与DNA类似，RNA含有腺嘌呤（A）、鸟嘌呤（G）和胞嘧啶碱基（C），但它含有尿嘧啶

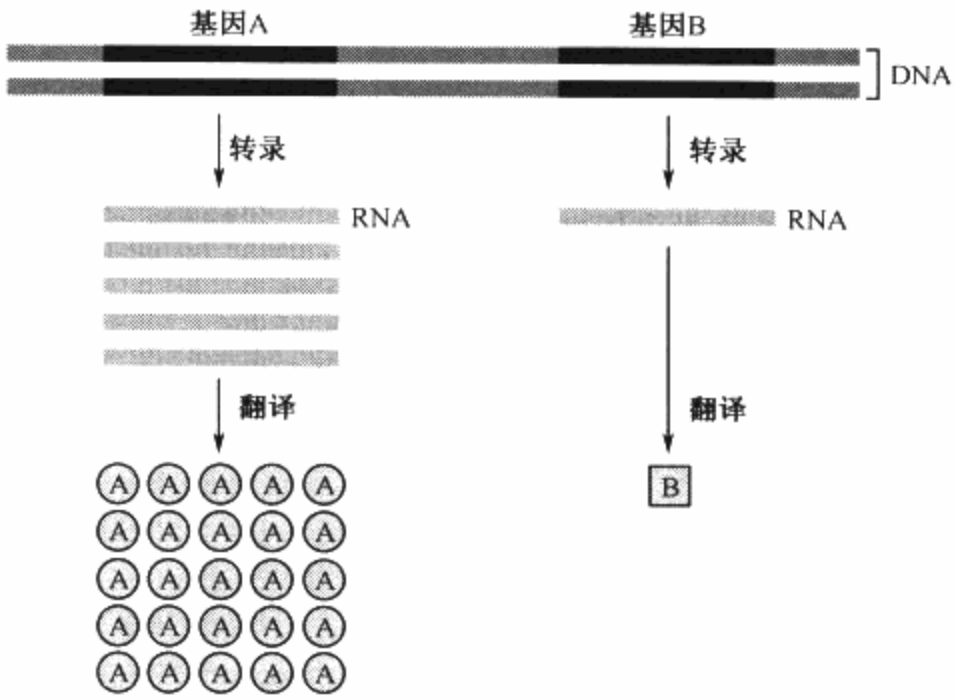


图 6-3　基因能以不同的效率表达

基因 A 转录和翻译比基因 B 更有效得多。这使得细胞中蛋白 A 的量比蛋白 B 的量大得多

(U)，而不是胸腺嘧啶（T)。由于 U 与 T 一样，能与 A 通过氢键形成碱基配对（图 6-5)，所以第 4 章和第 5 章中描述的 DNA 互补配对的特性也适用于 RNA（在 RNA 中，G 与 C 配对，A 与 U 配对)。然而，在 RNA 中也常常能找到其他罕见的碱基配对类型，如 G 偶尔与 U 配对。

除了这些化学上的微小差别以外，DNA 和 RNA 整体结构也显著不同。DNA 在细胞内总是以双链螺旋出现，RNA 却以单链出现。因此 RNA 折叠成不同的形状，就像一条多肽链折叠起来形成蛋白质的最终形状那样（图 6-6)。本章后面我们会看到，某些 RNA 分子折叠成复杂三维形状的能力赋予其结构和催化功能。

转录产生与 DNA 一条链互补的 RNA

细胞内所有的 RNA 都通过 DNA 转录产生，该过程与第 5 章讨论的 DNA 复制过程存在某些相似性。转录起始于 DNA 双螺旋一小部分打开并解旋，从而暴露出每条 DNA 链上的碱基。然后 DNA 双螺旋的两条链中有一条作为合成 RNA 分子的模板。与 DNA 复制一样，RNA 链的核苷酸序列由即将加入的核苷酸和 DNA 模板之间互补碱基配对决定。如果形成很好的配对，进入的核糖核苷酸就与生长的 RNA 链通过酶促反应共价连接起来。转录产生的 RNA 链（转录物）每次延长一个核苷酸，其核苷酸序列与用作模板的 DNA 链恰好互补（图 6-7)。

但是，转录在许多关键方面与 DNA 复制不同。与新合成的 DNA 链不同，RNA 链与 DNA 模板链之前并没有保持氢键连接。相反，就在加入核糖核苷酸的区域后面，RNA 链被取代，重新形成 DNA 螺旋。因此，转录产生的 RNA 分子以单链形式从 DNA 模板上释放下来。此外，RNA 分子比 DNA 分子短得多。一条人类染色体 DNA 分子可以长达 25 亿个核苷酸对：相比之下，大多数 RNA 长不过几千个核苷酸，而且许多还要短得多。

行使转录功能的酶称为 **RNA 聚合酶**。与催化 DNA 复制的 DNA 聚合酶类似（讨论见第 5 章)，RNA 聚合酶催化磷酸二酯键的形成，将核苷酸连接起来形成一条线性链。RNA 聚合酶沿 DNA 一步一步移动，在聚合作用活性位点前面将 DNA 螺旋解旋，以暴露出模板链的新区

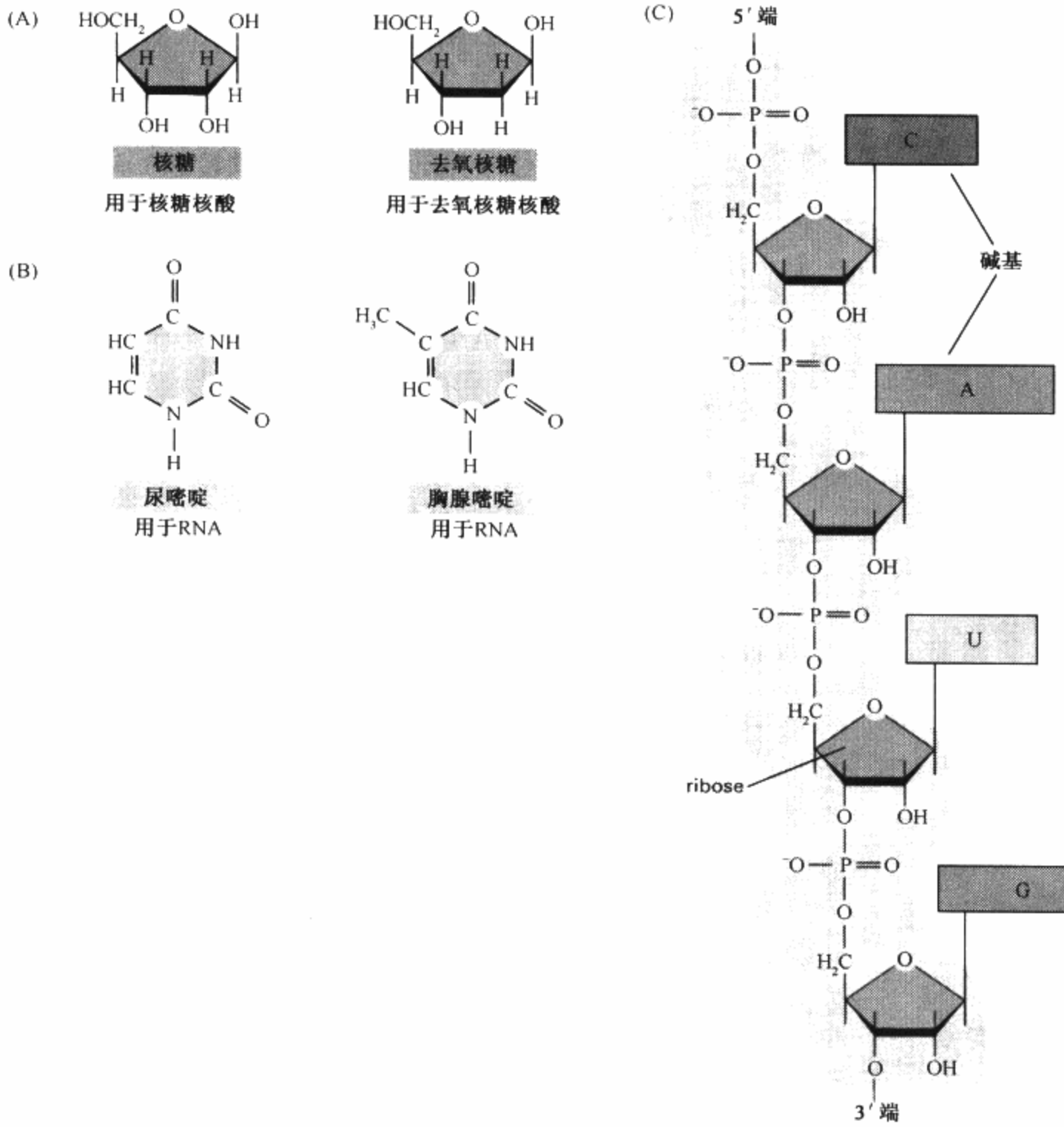


图 6-4　RNA 的化学结构

（A）RNA 含有核糖，它与 DNA 中使用的脱氧核糖不同，因为多一个额外的—OH 基团。（B）RNA 含有尿嘧啶碱基，它与 DNA 中的对等碱基，胸腺嘧啶不同，因为它缺少一个—CH_3基团。（C）一小段 RNA。RNA 中核苷酸之间的磷酸二酯化学键与 DNA 中的相同

域，供互补碱基配对。通过这种方式，生长的 RNA 链沿 5′→3′方向每次延伸一个核苷酸（图 6-8）。酶的底物是核苷三磷酸（ATP、CTP、UTP 和 GTP）；与 DNA 复制一样，高能键的水解提供能量推动反应向前进行（图 5-4）。

RNA 链在合成时几乎立即从 DNA 上释放，这意味着在第一条 RNA 转录完成之前，就能起始另一条 RNA 分子的合成，这样在相对较短的时间里可以从同一基因产生许多 RNA 拷贝（图 6-9）。通过这种方式，如果 RNA 聚合酶以 20 个核苷酸/s 的速度（真核生物中的速度），彼此紧跟着移动，那么 1h 内就能从单个基因合成出一千多条转录物。

虽然 RNA 聚合酶催化本质上与 DNA 聚合酶相同的化学反应，但是两种酶之间也存在一些重要的不同之处。首先，也是最明显的，RNA 聚合酶催化核糖核苷酸之间的连接，而不是

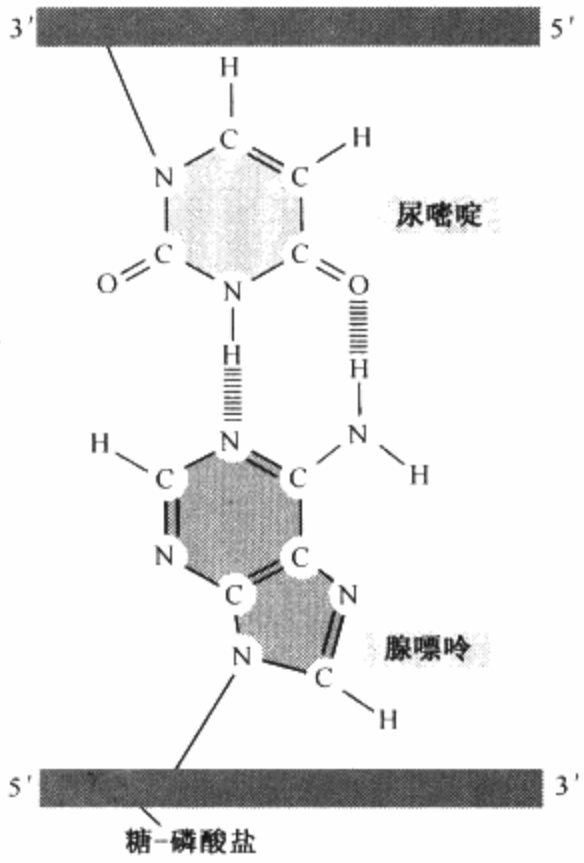


图 6-5　尿嘧啶与腺嘌呤配对

U 中缺少一个—CH_3 基团对碱基配对没有影响。U-A 碱基对与 T-A 碱基对非常相似（图 4-4）

脱氧核糖核苷酸。其次，与参与 DNA 复制的 DNA 聚合酶不同的是，RNA 聚合酶能够在没有引物的情况下起始 RNA 链合成。存在这种差别可能是因为转录不需要像 DNA 复制那样精确（表 5-1）。与 DNA 不同，在细胞中，RNA 并不永久性地储存信息。每拷贝 10^4 个核苷酸到 RNA 中，RNA 聚合酶就会出现 1 次错误（相比之下，DNA 聚合酶直接拷贝时错误率大约为每 10^7 个核苷酸 1 次），并且在 RNA 转录中出现错误的后果远不及 DNA 复制中那么明显。

虽然 RNA 聚合酶并不像复制 DNA 的 DNA 聚合酶一样精确，但它们具备忠实性有限的校对机制。如果错误的核糖核苷酸加入到生长的 RNA 链中，聚合酶就会倒退，酶的活性位点可以进行切除反应，除用的是 H_2O 而不是焦磷酸以外，它就像是聚合反应的逆反应（图 5-4）。RNA 聚合酶在错误整合的核苷酸附近停留时间，比它正确加入该核苷酸的时间要长，从而导致切除倾向于发生在错配核苷酸上。然而，作为提高精确性代价的一部分，RNA 聚合酶也切除许多正确的碱基。

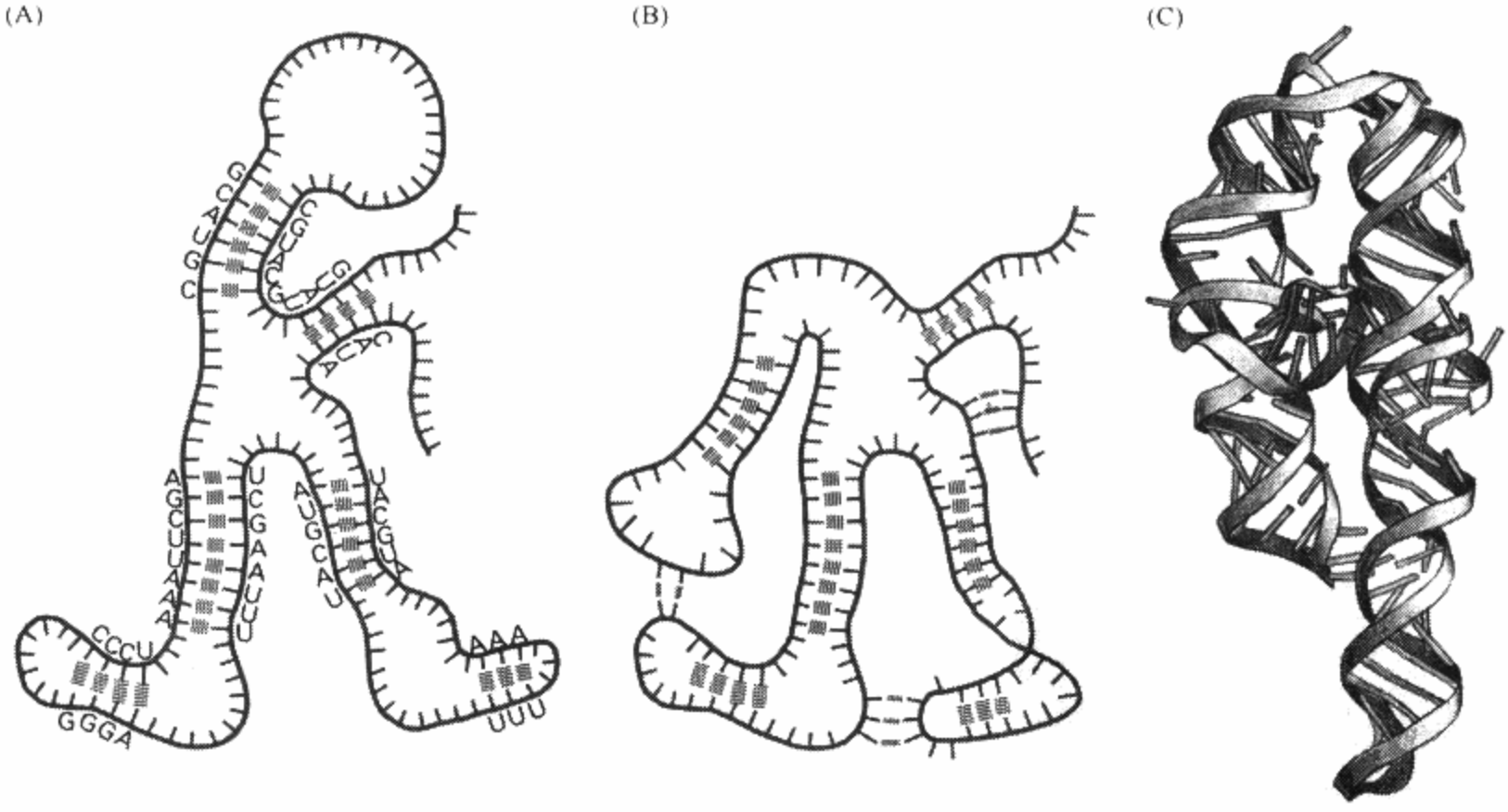


图 6-6　RNA 能折叠成特定的结构

RNA 大都是单链的，但它常常也含有小段核苷酸序列，能够与同一分子其他位置的互补序列形成常规碱基对。这些相互作用与其他“非常规”碱基对相互作用一起，使得 RNA 分子能折叠成由其核苷酸序列决定的三维结构。（A）只显示有常规碱基配对相互作用的折叠 RNA 分子图；（B）既有常规（红色），又有非常规（绿色）碱基配对相互作用的结构；（C）实际存在的 RNA 结构，类型Ⅰ内含子的一部分（图 6-36）。用双螺旋中的“横档”表示每个常规碱基配对相互作用。用断裂的横档表示以其他构型存在的碱基

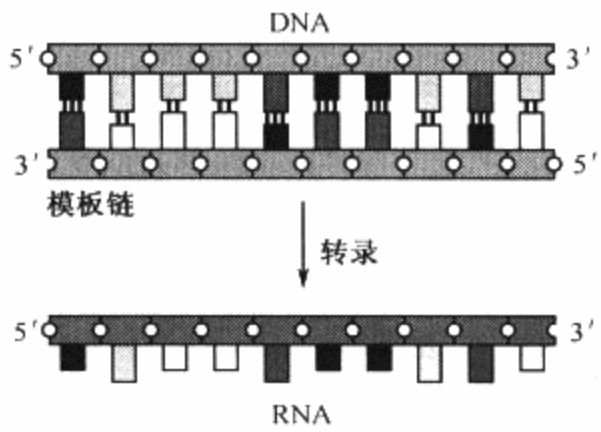


图 6-7　DNA 转录产生一条单链 RNA 分子，与其中一条 DNA 链互补

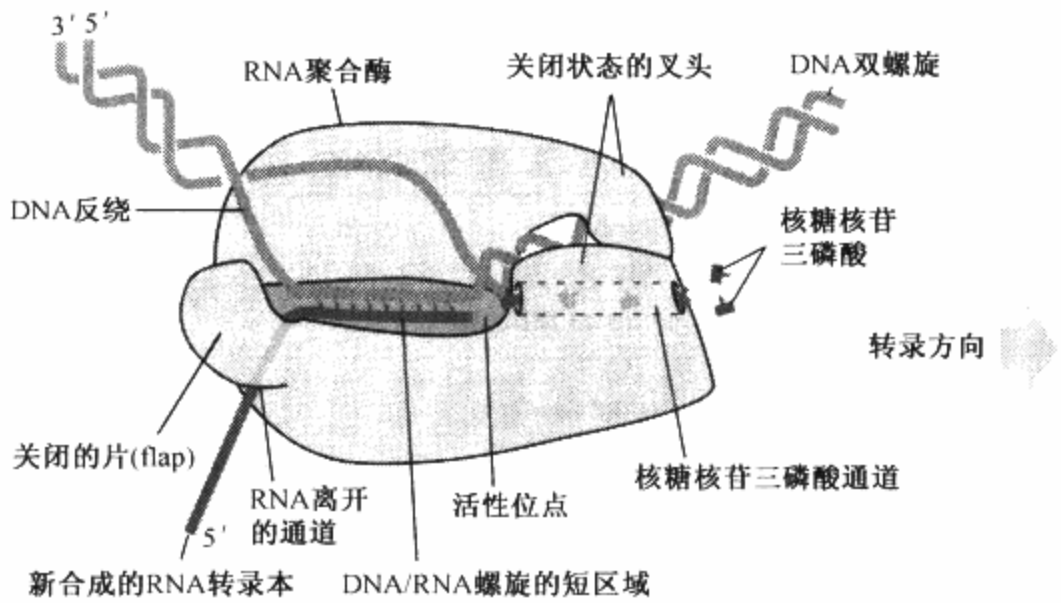


图 6-8　DNA 由 RNA 聚合酶进行转录

RNA 聚合酶（浅蓝色）沿 DNA 逐步移动，将 DNA 螺旋在其活性位点解旋。当它前进的时候，聚合酶利用一条裸露的 DNA 链作为模板，在聚合反应位点将核苷酸（这里用“T”形状代替）一个接一个地加到 RNA 链上。因而 RNA 转录物是两条 DNA 链其中一条的单链互补拷贝。聚合酶有一个舵（图 6-11），它置换出新合成的 RNA，从而使聚合酶后面的两条 DNA 链能够重新缠绕起来。因而只是暂时形成了一小段 DNA/RNA 螺旋区域（长将近 9 个核苷酸），所以一个 DNA/RNA 螺旋“窗户”随着聚合酶一起沿 DNA 移动。将要进入的核苷酸是核糖核苷三磷酸形式（ATP、UTP、CTP 和 GTP），储存在它们的磷酸-磷酸键中的能量提供了聚合反应的动力（图 5-4）（摘自 Robert Landick 友好提供的图）

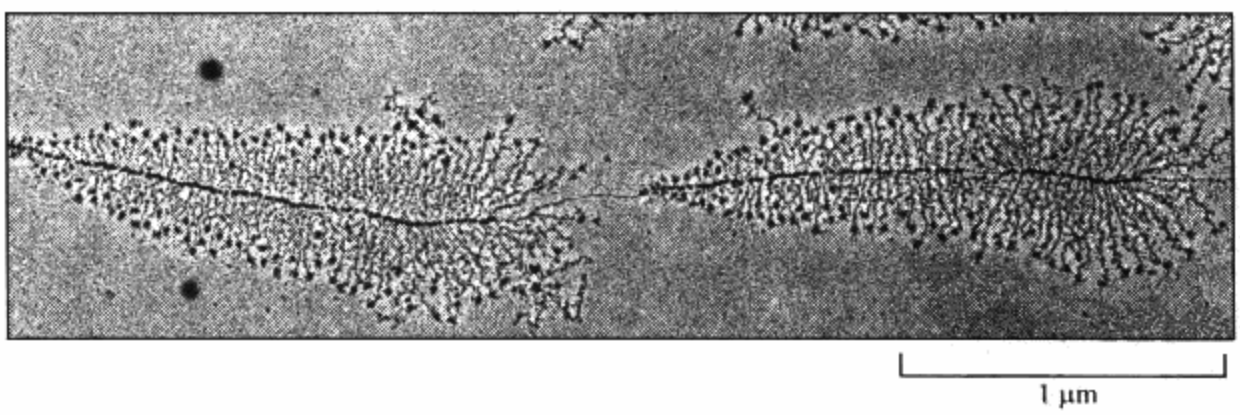


图 6-9　电镜下观察到的两个基因的转录

该显微照片显示出许多 RNA 聚合酶分子同时转录着两个相邻基因。可以看到 RNA 聚合酶分子以点状形式沿 DNA 排列的，新合成的转录物（细线）则附着在这些 DNA 上面。该例中显示的 RNA 分子（核糖体 RNA）并未翻译成蛋白质，而是直接用作核糖体的组分，它是进行翻译的装置。位于每个 rRNA 转录物 5′端（自由末端）的颗粒据相信反映了核糖体组装的开始。从新合成转录物的长度可以推出 RNA 聚合酶分子正从左向右进行转录（经 Ulrich Scheer 允许后使用）

细胞合成几种类型的 RNA

细胞 DNA 中携带的大部分基因指令了蛋白质的氨基酸序列；由这些基因拷贝得到的 RNA 分子（它们最终指导蛋白质的合成）称为信使 RNA（mRNA）分子。然而，少数基因的终产物却是 RNA 本身。仔细分析酿酒酵母基因组完整的 DNA 序列后发现，超过 750 个基因产生 RNA 作为终产物（一定程度上超过酵母基因总数的 10%），尽管这个数字包括了一些高度重复基因的大量拷贝。这些 RNA 与蛋白质一样，作为结构和功能组分在细胞内许多不同过程中发挥作用。第 5 章我们就遇到过其中一种 RNA，也就是由端粒酶携带的模板。虽然它们的功能并不全都清楚，但是本章我们会看到，许多核内小 RNA（snRNA）分子指导前体 mRNA 剪接成 mRNA，核糖体 RNA（rRNA）分子形成了核糖体的核心，还有转移 RNA（tRNA）分子形成了转换器，选择氨基酸并将其固定在核糖体上某个位置以便整合到蛋白质中（表 6-1）。

表 6-1　细胞内主要种类的 RNA

RNA 种类	功能
mRNA	信使 RNA，编码蛋白质
rRNA	核糖体 RNA，形成核糖体的，基本结构并催化蛋白质的合成
tRNA	转移 RNA，参与蛋白合成的主要 RNA，充当 mRNA 和氨基酸之间的转换器
snRNA	小核 RNA，在各种核过程中起发挥功能。包括前体 mRNA 的拼接
snoRNA	小核仁 RNA，加工和化学修饰 rRNAs
其他精编码 RNA	在不同的细胞过程中发挥功能，包括端粒合成，X 染色体失活和蛋白转运到内质网

每一段转录的 DNA 称为一个转录单位。在真核生物中一个典型的转录单位仅仅携带一个基因的信息。因而编码了或者一条 RNA 分子或者一个蛋白质（或者一组相关的蛋白质，如果最初的 RNA 转录物以不只一种方式剪接成不同的 mRNA）。在细菌中，一系列毗邻的基因通常作为一个单位转录，因此结果产生的 mRNA 分子携带了几个不同蛋白质的信息。

大体上讲，RNA 占细胞干物质重量的几个百分点。细胞内 RNA 大部分是 rRNA，在典型的哺乳动物细胞内，mRNA 只占总 RNA 的 3%～5%。mRNA 群体中含有上万种不同的分子，且平均每个细胞内每种 mRNA 只有 10～15 个分子存在。

编码在 DNA 中的信号告诉 RNA 聚合酶从哪里开始，又在哪里结束

为了精确地转录基因，RNA 聚合酶必须识别在基因组什么位置开始转录，又在什么位置结束转录。RNA 聚合酶完成这些工作的方式在细菌和真核生物之间有些差别。由于细菌中的过程简单一些，我们就先看看在其中如何进行的。

转录起始是基因表达中特别重要的一步，因为它是细胞调节合成哪种蛋白质，以什么速率合成的主要点。细菌 RNA 聚合酶是一个多亚基复合物。一个可解离的亚基，称为 σ 因子，主要负责 RNA 聚合酶阅读 DNA 中告诉它从哪里起始转录的信号的能力（图 6-10）。RNA 聚合酶分子碰到细菌 DNA 以后，只能微弱地附着在上面，每个酶分子通常迅速沿 DNA 长分子滑动，直到再次解离为止。然而，当聚合酶滑到 DNA 双螺旋上一段称为**启动子**的区域，也

就是表明 RNA 合成开始位点的一段特殊核苷酸序列以后，它就牢固地结合在上面。聚合酶使用它的 σ 因子，通过与裸露在螺旋外的碱基部分产生特异性相互作用，识别这段 DNA 序列（图 6-10 中步骤 1）。

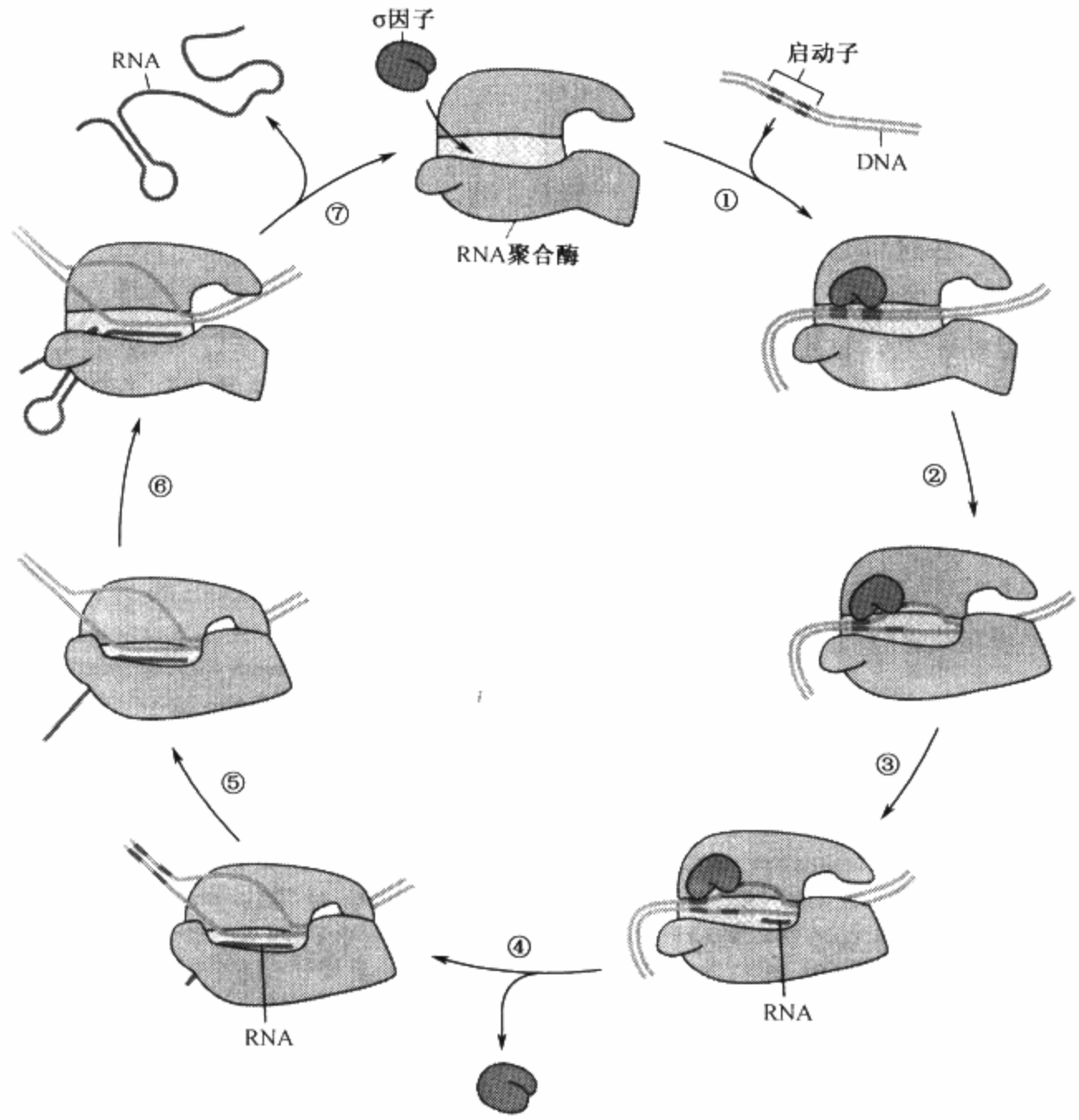


图 6-10　细菌 RNA 聚合酶的转录循环

步骤 1 中，RNA 聚合酶全酶（核心酶加上 δ 因子）形成，然后定位到启动子上（图 6-12）。聚合酶在转录开始的地方将 DNA 解旋（步骤 2）并开始转录（步骤 3）。这个最初的 RNA 合成（有时也称作“流产起始”）效率比较低。然而，一旦 RNA 聚合酶已经合成出 10 个核苷酸的 RNA，δ 因子就不再紧密结合，聚合酶发生一系列构象变化（这些变化可能包括收紧钳口和将 RNA 放到出口通道（图6-11）。现在聚合酶转变成合成 RNA 的延伸模式（步骤 4），本图中它沿着 DNA 向右侧移动。在延长模式期间（步骤 5）转录是高度持续进行的，只有遇到终止信号（步骤 6）时聚合酶才会离开 DNA 模板并释放新转录的 RNA。终止信号编码在 DNA 中，许多都通过形成一个 RNA 结构，使聚合酶对 RNA 的维系失去稳定，如图所示。在细菌中，所有 RNA 分子都由同一类 RNA 聚合酶合成，因而图中描述的循环适用于 mRNA 合成以及起结构和催化作用 RNA 的合成（摘自 Robert Landick 友好提供的图）

RNA 聚合酶以这种方式牢固地结合到启动子 DNA 上以后，它就打开双螺旋，裸露出两条链上的一小段核苷酸（图 6-10 中步骤 2）。与 DNA 解螺旋酶催化的反应不同（图 5-15），打开有限的一段螺旋不需要 ATP 水解的能量。相反，聚合酶和 DNA 都发生可逆的结构变化，使它们处于能量上更加有利的状态。随着 DNA 的解旋，两条裸露 DNA 链中有一条作为

与即将加入的核糖核酸互补配对的模板（图 6-7），为了起始 RNA 链的合成，这些核苷酸中有两个是通过聚合酶连接起来。最初 10 个左右核苷酸的 RNA 合成出来以后（这是一个效率相对较低的过程，其间 RNA 聚合酶合成并丢弃短核苷酸寡聚物），σ 因子松弛它与聚合酶的紧密结合，并最终从酶上面解离下来。在这个过程期间．聚合酶发生进一步的结构变化，使它能够迅速向前移动，从而在不存在 σ 因子的情况下进行转录（图 6-10 中的步骤 4）。链不断延长（对于细菌 RNA 聚合酶，以将近 50 个核苷酸/s 的速度）直到酶遇上 DNA 中另一种信号——**终止子**（下面描述），即聚合酶停止并释放出 DNA 模板和新合成的 RNA 链的位置（图 6-10 中步骤 7）。聚合酶从终止子位点上释放以后，它与一个游离的 σ 因子重新结合，并搜寻新的启动子，在这里它能再次起始转录过程。

细菌 RNA 聚合酶许多结构特征使其特别擅长进行刚才描述的转录循环。一旦 σ 因子将聚合酶定位到启动子上，且模板 DNA 已经解旋，并推进到活性位点，就认为它像一对可移动的颚一样夹在 DNA 上（图 6-11）。当转录出前 10 个核苷酸后，σ 因子的解离使得聚合酶背部的口盖封闭起来形成一个出口通道，让新合成的 RNA 离开酶。此时，聚合酶以它的延伸模式工作时，酶上一个像舵一样的结构不断地将形成的 DNA-RNA 杂交分子撬开。我们可以将转录起始期间发生的一系列构象改变，看成是酶围绕 DNA 和 RNA 接连的收紧过程，以确保在转录完一个基因之前，酶不会解离：如果 RNA 聚合酶确实提前解离了，它就不能继续延伸，而是必须在启动子上重新开始转录。

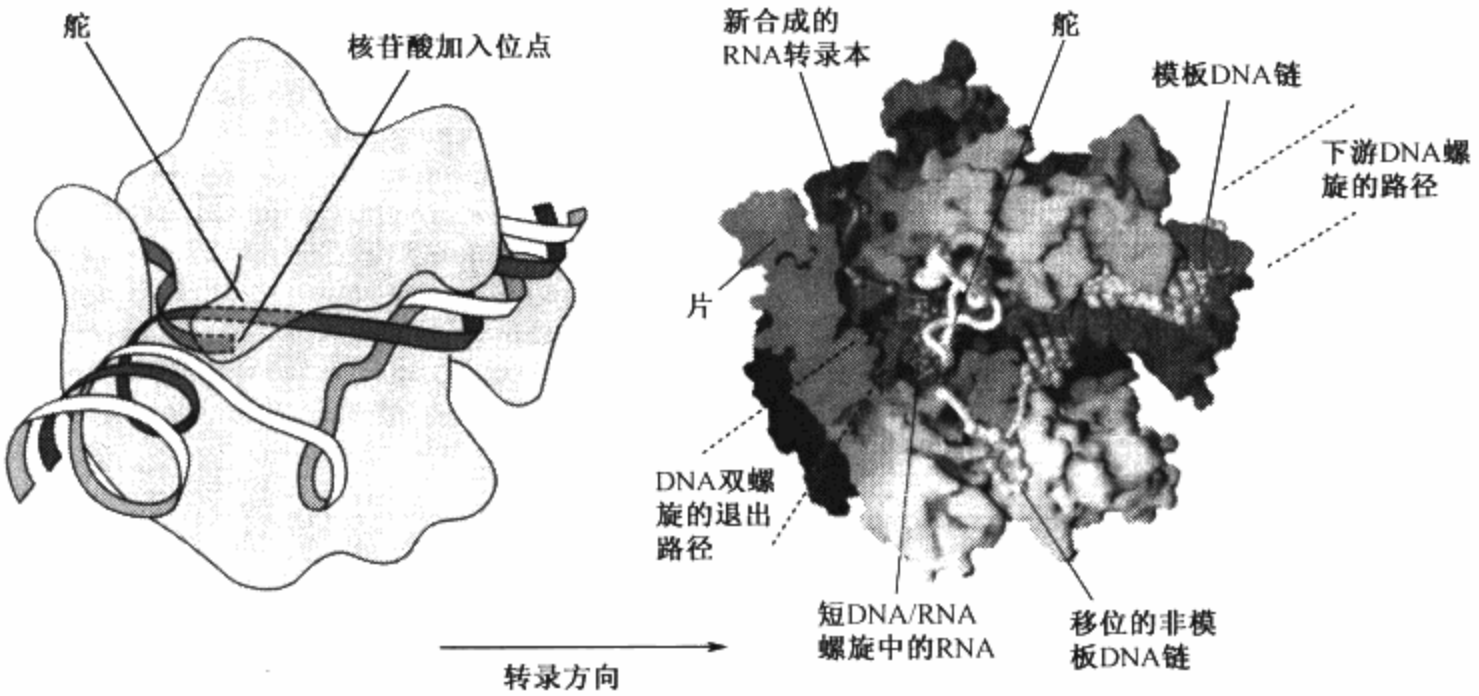


图 6-11　细菌 RNA 聚合酶的结构

两幅描绘细菌 RNA 聚合酶三维结构的图，其中含有 DNA 和 RNA 作为模型。该 RNA 聚合酶由 4 个不同亚基组成，它们分别用不同的颜色示出（右边）。红色的是用作模板的 DNA 链，黄色的是非模板链。当聚合酶移动时，方向舵将 DNA-RNA 杂交分子挤开。为简单起见，右边的图中只显示了方向舵的骨架，从聚合酶出来的 DNA 也已略去。由于描绘的 RNA 聚合酶处于延长模式，所以没有 δ 因子（经 Seth Darst 允许后使用）

DNA 中的信号（终止信号）如何停止正在延伸的聚合酶呢？对于大多数细菌基因，终止信号由一串 A-T 核苷酸对跟随一段两重对称的 DNA 序列组成，这段对称序列一旦转录成 RNA，就会通过 Watson-Crick 碱基配对折叠成一个“发卡”结构（图 6-10）。当聚合酶转录通过终止子时，这个发卡拷贝有助于将 RNA 聚合酶上可移动的盖子抵住，并将 RNA 转录物从出口通道释放出来。同时活性位点上的 DNA-RNA 杂交分子由于主要通过 U-A 碱基对维系（它不如 G-C 碱基对稳定，因为每个碱基对之间只形成 2 个氢键而不是 3 个），不够牢固，

难以将 RNA 保持原位，所以发生解离，可能是通过强行打开了聚合酶，导致它从 DNA 上释放下来。由此可见，在有些方面，转录终止似乎涉及起始时结构转变的逆向进行。转录终止过程也是本章普通主题之一：RNA 折叠成特殊结构的能力在解读基因组过程的许多方面发挥了主要作用。

转录起始和转录终止信号是不同的核苷酸序列

我们刚刚看到，转录起始和终止这两个过程涉及蛋白质，DNA 和 RNA 中一系列复杂的结构转变。研究者很难识别编码在 DNA 中指令这种转变的信号，可能也并不奇怪。实际上，比较许多不同细菌启动子后揭示出它们在 DNA 序列上是不同类的。然而，它们都含有相关的序列，在某些方面反映了直接由 σ 因子识别的 DNA。常常以共有序列的形式概括这些共同特征（图 6-12）。一般来说，共有核苷酸序列是通过比较许多具有相同基本功能的序列，并统计每个位点上最常见核苷酸之后得到。因此它可以作为大量的单个核苷酸序列的概括或者“平均”。

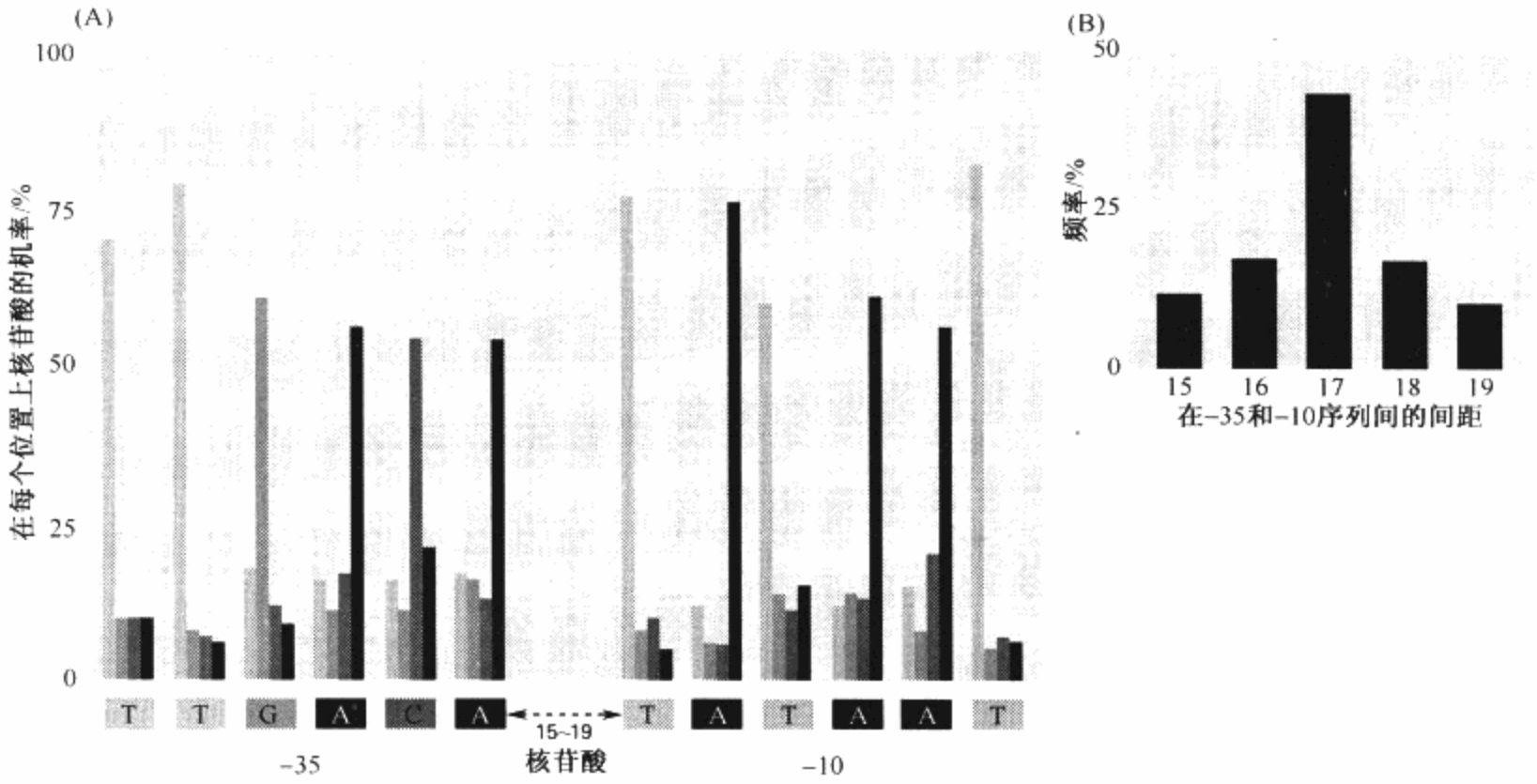


图 6-12　大肠杆菌主要类型启动子的共有序列

（A）启动子特征性地含有 2 个六聚体 DNA 序列，它们是－35 序列和－10 序列，分别以它们距转录起点（指定其为＋1)的相对位置命名。为了方便，显示了一条单链 DNA 的核苷酸序列；实际上 RNA 聚合酶识别双链 DNA 的启动子。在比较了 300 个启动子的基础上，给出了 4 个核苷酸在－35 和－10 六聚体中各个位置的频率。图下方显示的共有序列反映了一组启动子中各个位置发现的最常见的核苷酸。位于－10 和－35 六聚体之间的核苷酸序列在这些启动子中间没有表现出重要的相似性。（B）大肠杆菌启动子中发现的－35 六聚体和－10 六聚体之间间隔的分布。这两幅图显示的信息适用于 RNA 聚合酶识别的大肠杆菌启动子和主要的 δ 因子（命名为 δ^{70}）。下一章我们将看到，细菌也含有较次要的 δ 因子，它们每个都识别一段不同的启动子序列。有些由 RNA 聚合酶和 δ^{70} 识别的特别强的启动子拥有其他的序列，它们位于－35 六聚体上游（图中向左），并由 RNA 聚合酶的另一个亚基识别

各个细菌启动子 DNA 序列不相同的原因之一在于精确的序列决定了启动子的强弱（或者每单位时间起始转录的次数）。进化过程因此已对每个启动子进行了调试，以便它们能够尽可能地频繁起始转录，此外，进化过程已形成了大量不同的启动子。编码高丰度蛋白质的基因启动子比那些编码稀有蛋白质的启动子要强得多，正是它们的核苷酸序列产生

了这些差别。

与细菌启动子一样，转录终止子也包括广泛不同的序列，它们重要的共同特征就是有可能形成简单的 RNA 结构。由于几乎有无限多的核苷酸序列具备这种潜能，所以终止子序列与启动子序列相比更加不同。

我们已经比较详细地讨论了细菌启动子和终止子，以便说明与基因组序列分析有关的一个重要方面。虽然我们对细菌启动子和终止子了解很清楚，而且能找出概括它们最突出特征的共有序列，但是由于它们核苷酸序列的变化，使得研究者们（即便有功能强大的电脑帮助）很难仅仅通过审查基因组序列就明确地将它们定位。如果我们在真核生物中遇到相似类型的序列，想要定位它们就更加难。通常需要其他的信息，来精确地定位出基因组中含有的短 DNA 信号，其中一些信息就直接来自实验。

启动子序列是不对称的（图 6-12）。这个特征对于它们在染色体中的排列产生重要影响。由于 DNA 是双链的，从理论上讲，利用两条 DNA 链中每一条作为模板，可以从任何一个基因转录得到的两条不同的 RNA 分子。然而一个基因通常只有一个启动子，并且由于细菌启动子核苷酸序列是不对称的，因此聚合酶只能以一个方向结合。可见聚合酶没有选择余地，而是只能转录其中一条 DNA 链，因为它只能以 5′→3′方向合成 RNA（图 6-13）。每个基因模板链的选择因而由启动子的位置和方向决定。基因组序列揭示了不同的基因中用作 RNA 合成模板的 DNA 链是不同的（图 6-14，也可见图 1-31）。

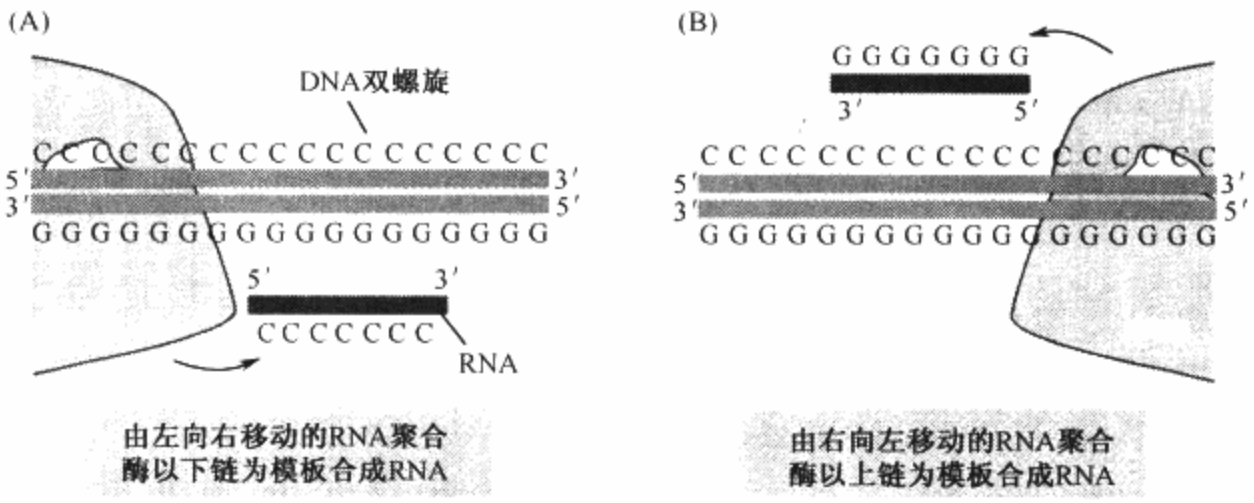


图 6-13　RNA 聚合酶定向的重要性

作为模板的 DNA 链必须以 3′→5′方向进行跨越，如图 6-9 所示。因此，RNA 聚合酶移动的方向决定了两条 DNA 链中哪一条作为模板合成 RNA，如（A）和（B）中所示。聚合酶方向继而又由启动子序列，即 RNA 聚合酶开始转录的位点方向所决定

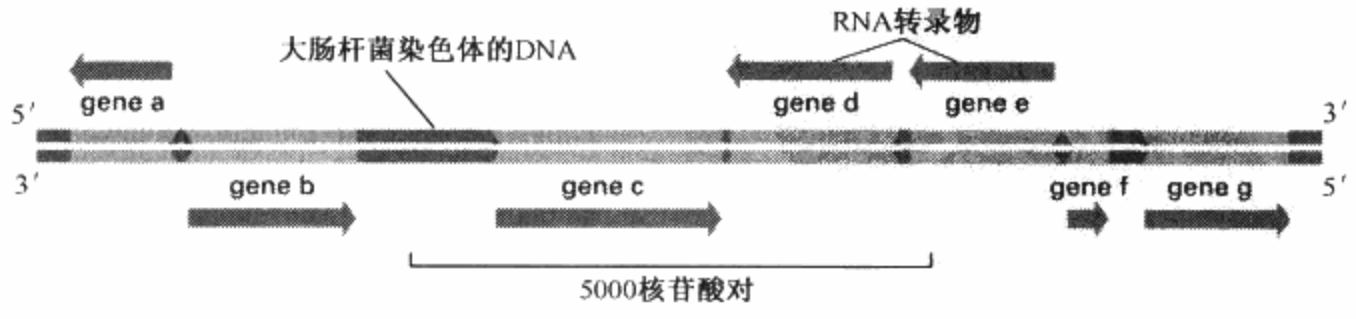


图 6-14　沿一小部分细菌染色体进行转录的方向

有些基因利用一条 DNA 链作为模板转录，而另一些基因又利用另外一条 DNA 链作为模板转录。转录方向由每个基因起始处的启动子决定（绿色箭头）。这里描绘了大肠杆菌染色体的将近 0.2%（9000 个碱基对）。基因转录利用下面的 DNA 链作为模板从左向右进行；那些从右向左转录的基因利用上面的链作为模板

考查了细菌中的转录以后，我们现在看看真核生物中的情况，其 RNA 分子的合成是一个更加复杂的事情。

真核生物中转录起始需要许多蛋白质

与含有一种 RNA 聚合酶的细菌相比，真核生物细胞核存在 3 种 RNA 聚合酶，分别称为 RNA 聚合酶Ⅰ、RNA 聚合酶Ⅱ和 RNA 聚合酶Ⅲ。这 3 种聚合酶结构上彼此相似（也与细菌 RNA 聚合酶相似）。它们拥有一些共同的亚基和许多结构特征，但它们转录不同类型的基因（表 6-2）。RNA 聚合酶Ⅰ和Ⅲ转录转移 RNA、核糖体 RNA 和各种小 RNA 的编码基因；RNA 聚合酶Ⅱ转录大多数基因，包括那些编码蛋白质的基因，因此我们接着主要集中对这个酶进行讨论。

表 6-2　真核细胞中三种 RNA 聚合酶

聚合酶的类型	转录的基因
RNA 聚合酶Ⅰ	5.8S,18S 和 28S rRNA 基因
RNA 聚合酶Ⅱ	全部蛋白编码的基因,加上 snoRNA 基因和一些 snRNA 基因
RNA 聚合酶Ⅲ	tRNA 基因,5S rRNA 基因,一些 snRNA 基因和来自其他小 RNAs 的基因

虽然真核生物 RNA 聚合酶Ⅱ与细菌 RNA 聚合酶有许多结构相似之处（图 6-15），但是在细菌和真核生物中的这种酶发挥功能的方式上却存在许多重要的差别，现在我们关注一下其中的两点。

图 6-15　细菌 RNA 聚合酶和真核生物 RNA 聚合酶Ⅱ的结构相似性

绿色表示这两种 RNA 聚合酶结构相似的区域。真核生物聚合酶比细菌聚合酶大（12 个亚基而不是 5 个亚基），灰色表示一些其他的区域。蓝色球代表作为聚合酶结构组分的 Zn 原子，红色球代表存在于发生聚合反应的活性位点的 Mg 原子。所有现今细胞（细菌、古细菌和真核生物）中的 RNA 聚合酶都密切相关，表明在生命的这三种主要分支趋异以前该酶的基本特征就已经具备了（经 P. Cramer 和 R. Kornberg 允许后使用）

1. 虽然细菌 RNA 聚合酶在体外没有其他蛋白质帮助时（具有 σ 因子作为其中一个亚基）能够在 DNA 模板上起始转录。但是真核 RNA 聚合酶不能，它们需要大量称作通用转录因子

的蛋白质的帮助，在聚合酶能开始转录之前这些蛋白质必须和聚合酶在启动子上组装。

2. 真核生物起始必须解决 DNA 包装成核小体以及更高级染色体结构形式这些问题，而细菌染色体却不具备这些特征。

RNA 聚合酶Ⅱ需要通用转录因子

与细菌 RNA 聚合不同，纯化的真核生物 RNA 聚合酶Ⅱ不能在体外起始转录，这个发现导致找到并纯化了起始过程所需的其他因子。这些通用转录因子帮助将 RNA 聚合酶正确地定位在启动子上，也有助于将 DNA 两条链分开，使转录能够开始，并在转录一旦起始以后，将 RNA 聚合酶从启动子释放到延伸模式。这些蛋白质之所以是“通用”的，是因为它们在 RNA 聚合酶Ⅱ利用的所有启动子上组装；它们由一系列相互作用的蛋白质组成，并命名为 TFⅡ（指聚合酶Ⅱ的转录因子），其中有 TFⅡA 和 TFⅡB 等。从广义上讲，真核生物通用转录因子发挥与细菌 σ 因子相同的功能。

图 6-16 显示了通用转录因子怎样在体外组装到 RNA 聚合酶Ⅱ利用的启动子上。组装过程开始于通用转录因子 TFⅡD 结合到一段主要由 A 和 T 组成的双螺旋 DNA 短序列上。因为如此，这个序列称为 TATA 序列，或者 **TATA 框**。识别该序列的 TFⅡD 的亚基称为 TBP（TATA 结合蛋白）。TATA 框通常位于转录起点上游 25 个核苷酸的位置。它并不是唯一的为转录开始提供信号的 DNA 序列（图 6-17），但对于大部分聚合酶Ⅱ启动子而言，它是最重要的。TFⅡD 的结合引起 TATA 框中的 DNA 产生很大扭曲（图 6-18）。这种扭曲被认为是一种物理标志，表明一个活性启动子在巨大的基因组中的位置，并且它将扭曲两侧的 DNA 序列带到一起，使下面的蛋白质组装过程能够进行。然后其他的因子与 RNA 聚合酶Ⅱ一起组装起来，形成一个完整的转录起始复合物（图 6-16）。

RNA 聚合酶Ⅱ被引导至启动子 DNA 上并形成了转录起始复合物以后，必须在转录起始位点接触到模板链。这一步骤得到一种通用转录因子 TFⅡH 的帮助，它含有一个 DNA 解旋酶。下一步，与细菌聚合酶一样，聚合酶Ⅱ停留在启动子位置，合成了一小段 RNA，直到它发生了构象变化并得以释放，才开始转录基因。在这个释放过程中关键的一步就是在 RNA 聚合酶“尾巴”（也称作 CTD 或 C 端结构域）上增加一个磷酸基团。整个磷酸化过程也由 TFⅡH 催化，除了解旋酶以外，它的一个亚基还是一种蛋白质激酶[图6-16(D)和(E)]。然后聚合酶可以从这串通用转录因子中脱离出来，同时发生一系列构象变化以加强它与 DNA 的相互作用，并募集新蛋白质以使它能够转录很长一段距离而不会解离。

一旦聚合酶Ⅱ开始延长 RNA 转录物，大部分通用转录因子就会从 DNA 上释放，这样它们就能用来与一个新的 RNA 聚合酶分子一起启动下一轮转录。很快我们就会看到，RNA 聚合酶Ⅱ“尾巴”的磷酸化也引起 RNA 加工装置的组分结合到聚合酶上，因而得以就位并在新转录的 RNA 从聚合酶刚出来就对它进行修饰。

聚合酶Ⅱ也需要激活蛋白、介导蛋白以及染色质修饰蛋白

以上描述的转录起始模型是在体外对 RNA 聚合酶Ⅱ和它的通用转录因子在纯化的 DNA 模板上的活性进行研究后建立起来的。然而，正如第 4 章讨论的那样，真核细胞中的 DNA 包装成核小体，然后这些核小体进一步组装成更高级的染色质结构。结果，真核细胞内转录起始过程比在纯化 DNA 上的过程更加复杂。首先，称作转录激活蛋白的基因调节蛋白质结合到 DNA 中特定的序列上，帮助吸引 RNA 聚合酶Ⅱ到转录起始位点（图 6-19）。RNA 聚合酶和通用转录因子需要这种吸引作用，以便帮助自身克服结合包装成染色质的 DNA 的困难。

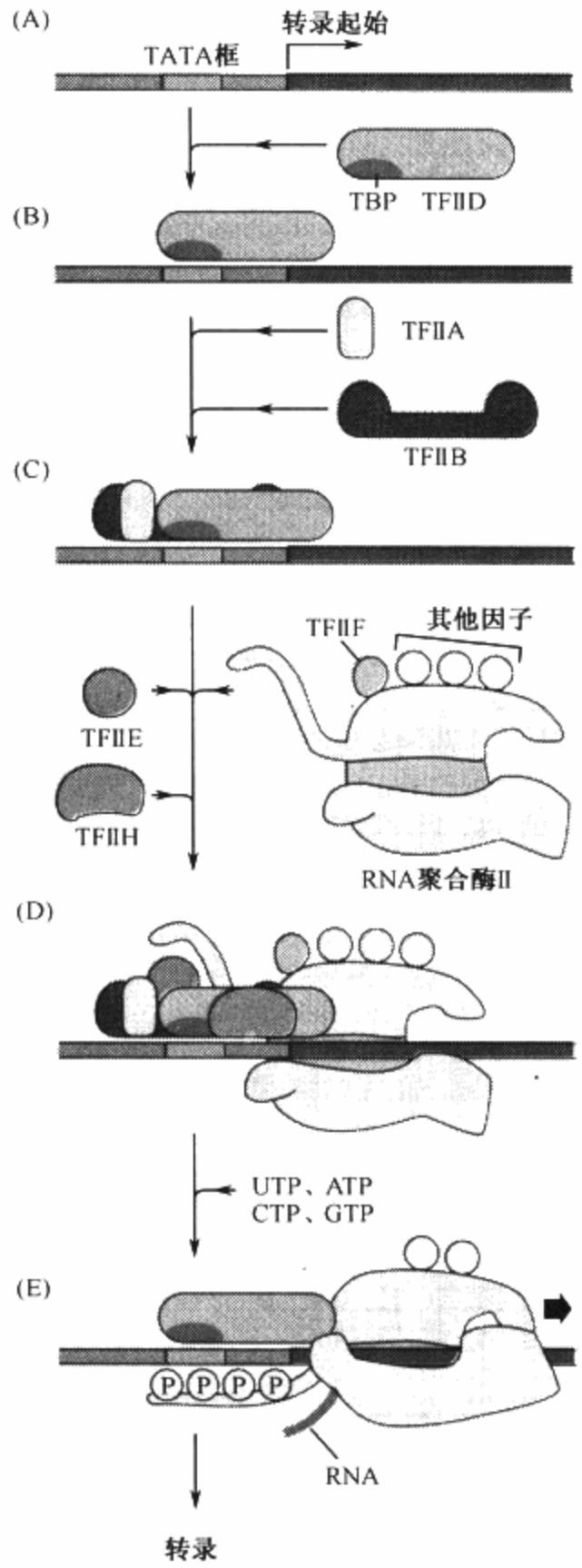
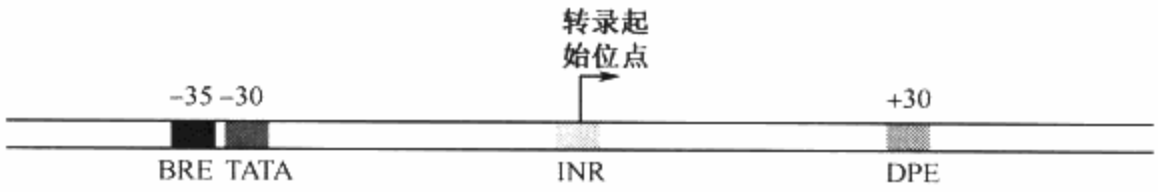


图 6-16　RNA 聚合酶Ⅱ催化的真核基因转录起始

RNA 聚合酶要开始转录就需要许多通用转录因子（称为 TFⅡA、TFⅡB 等）。(A) 启动子含有一段称为 TATA 框的 DNA 序列，它位于离转录起始位点 25 个核苷酸处。(B) TATA 框由转录因子 TFⅡD 识别并结合，然后该因子使 TFⅡB 能够结合到邻近位置。(C) 为了简便起见未显示出 TFⅡD 结合引起的 DNA 扭曲（图 6-18）。(D) 其余的通用转录因子，以及 RNA 聚合酶本身，在启动子上组装起来。(E) TFⅡA 然后利用 ATP 撬开转录起始位点的 DNA 双螺旋，使转录得以开始。TFⅡH 也磷酸化 RNA 聚合酶Ⅱ，从而改变其形状，这样聚合酶就从通用转录因子上释放，并能够开始转录的延伸阶段。如图所示，磷酸化的位点是一个从聚合酶分子伸出来的 C 端长多肽尾。图中显示的组装简图从体外进行的实验中推测而来，在细胞中通用转录因子组装到启动子上的实际顺序还不确定。在有些情况下，认为通用转录因子首先与聚合酶组装，随后整个组装物经过一步结合到 DNA 上。通用转录因子在进化中高度保守；通过生物化学实验，有些来自人类细胞的通用转录因子能用来自简单酵母中的对应转录因子替换



元件	共同序列	通用转录因子
BRE	G/C G/C G/A C G C C	TFIIB
TATA	T A T A A/T A A/T	TBP
INR	C/T C/T A N T/A C/T C/T	TFIID
DPE	A/G G A/T C G T G	TFIID

图 6-17　真核生物 RNA 聚合酶Ⅱ起始位点附近发现的共有序列

图中显示了每个共有序列（第一排）和识别它的通用转录因子的名称（最后一排）。N 表示任何核苷酸，由斜线分隔的两个核苷酸表示在所示位置任何一个核苷酸出现的可能性相同。实际上，每个共有序列都是与图 6-12 类似的直方图的简略表示。对于大多数 RNA 聚合酶Ⅱ转录起始位点，只存在这 4 种序列中的两种或三种。例如，大多数 RNA 聚合酶Ⅱ启动子拥有一个 TATA 框序列和那些通常没有“强”INR 的序列。虽然大多数影响转录起始的 DNA 序列位于转录起始位点“上游”，但也有一些例外，如该图中显示的 DPE，位于转录的区域中间

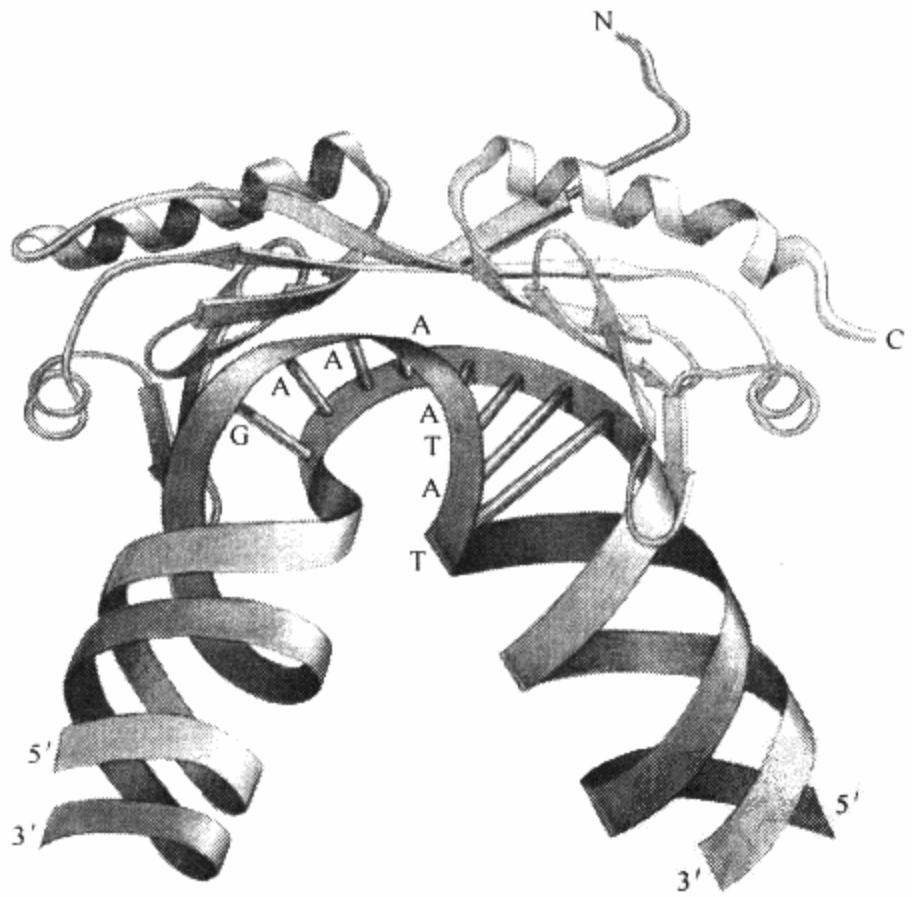


图 6-18　结合着 DNA 的 TBP（TATA 结合蛋白）的三维结构

TBP 是通用转录因子 TFⅡD 的一个亚基，后者负责识别并结合 DNA 中的 TATA 框序列（红色）。TBP 引起的独特 DNA 弯曲（部分解旋的 DNA 将双螺旋中的两个结节分隔开），可能作为标记，帮助吸引其他的通用转录因子。TBP 是一条折叠成两个非常相似结构域的多肽链（蓝色和绿色）（摘自 J. L. Kim et al., *Nature* 365：520～527，1993）

我们将在第 7 章讨论激活蛋白的作用，因为它们代表细胞调节基因表达的主要途径之一。这里我们只需注意真核细胞内转录的起始需要一种称为介导蛋白的蛋白质存在，它使激活蛋白能与聚合酶Ⅱ以及通用转录因子进行正确的相互作用。最后，细胞内转录起始常常需要在局

部募集染色质修饰酶，包括染色体重建复合体和组蛋白乙酰酶（图 6-19）。第 4 章讨论过，这两种酶能帮助更好地接触染色质 DNA，这样的话它们就会促进转录起始装置组装到 DNA 上。

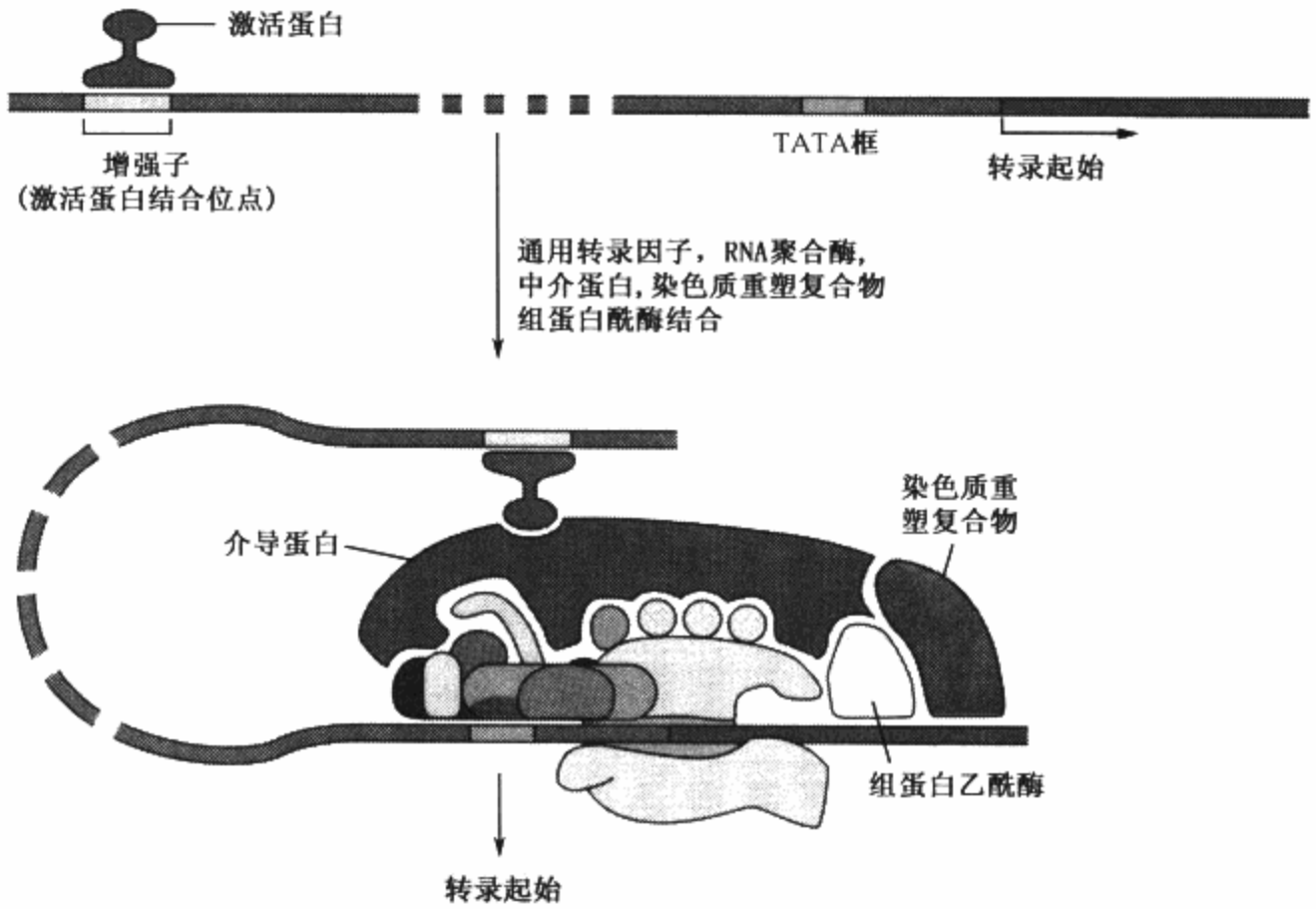


图 6-19　真核细胞中 RNA 聚合酶Ⅱ催化的转录起始

体内转录起始需要存在转录激活蛋白。如第 7 章所述，这些蛋白质结合 DNA 中特定的短序列。虽然这里只显示了一个，但是典型的真核细胞有许多激活蛋白，它们共同决定其转录的速率和模式。有时这些基因调节蛋白从远离几千核苷酸对的距离发挥作用（用虚线 DNA 分子表示），它们帮助 RNA 聚合酶，通用转录因子和介导因子在启动子上组装。此外，激活蛋白吸引依赖于 ATP 的染色质重建复合体和组蛋白乙酰酶。第 4 章曾讨论过，染色质的默认状态可能是 30nm 纤丝，并且这可能是转录发生起始的 DNA 形式。为简明起见，图中未示出

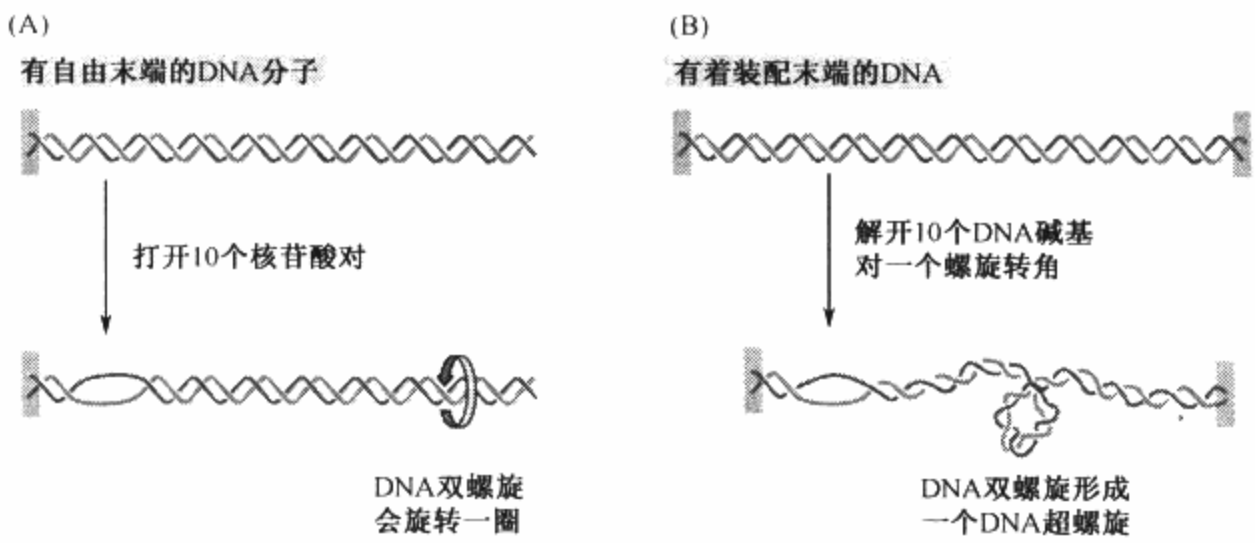
如图 6-19 所示，为了在真核细胞内起始转录，许多蛋白质（一百多种单个亚基）必须组装到转录起始位点。这些蛋白质的组装顺序对于不同的基因可能不同，因而并不是沿着规定好的途径进行的。实际上，这些不同的蛋白质组装物中有些可能远离 DNA 就相互作用，并作为预先形成的亚复合体被携带到 DNA。例如，中介蛋白、RNA 聚合酶Ⅱ，以及有些通用转录因子，能在内质网中彼此结合，并作为一个单位被带至 DNA。第 7 章我们再回到这个问题上来，其中我们会讨论真核细胞调节转录起始过程的许多方式。

转录延伸在 DNA 中产生超螺旋张力

RNA 聚合酶一旦起始了转录，并不会平稳地沿 DNA 分子前进；相反，它会颠簸着移动，在有些序列上短暂停留，也会迅速转录通过另外一些序列。延伸中的 RNA 聚合酶，无论是在细菌还是在真核生物中，都和一系列延伸因子结合，它们是降低 RNA 聚合酶到达基因终点之前解离可能性的蛋白质。这些因子通常在起始发生后不久就与聚合酶结合，并帮助 RNA 聚合酶移动通过基因中发现的各种 DNA 序列。真核 RNA 聚合酶沿 DNA 模板移动的时候，也必须与染色体结构竞争。实验表明，细菌中的聚合酶永远不会在体内遇到核小体，然

而能在体外转录通过它们，这说明核小体是容易跨越的。然而，真核生物聚合酶必须移动通过比简单的核小体更加紧密的染色质结构。因此看起来它可能在染色质重建复合体的帮助下进行转录。这些复合体可能与聚合酶一起移动，或者可能只是找到并解救停下来的聚合酶。此外，一些与 RNA 聚合酶结合的延伸因子促进转录在不需额外能量的情况下通过核小体。整个过程是如何完成的还不清楚，但是聚合酶转录核小体 DNA 时，这些蛋白质可能帮助挤出核小体核心组分。

对于延伸的聚合酶来说，仍然存在另一个障碍，无论是在细菌还是在真核生物中。讨论这个问题时，我们需要先考虑 DNA 双螺旋固有的一种微妙特性，称为 **DNA 超螺旋**。DNA 超螺旋代表了 DNA 为应对超螺旋张力而采用的一种构象；在螺旋中形成环或者卷曲就会形成这种张力，该过程可逆。图 6-20(A)说明了想像的一种引起 DNA 超螺旋的拓扑限制的简单方式。在 DNA 双螺旋中，每圈螺旋大概有 10 个核苷酸对。想像一个两端对彼此而言都固定



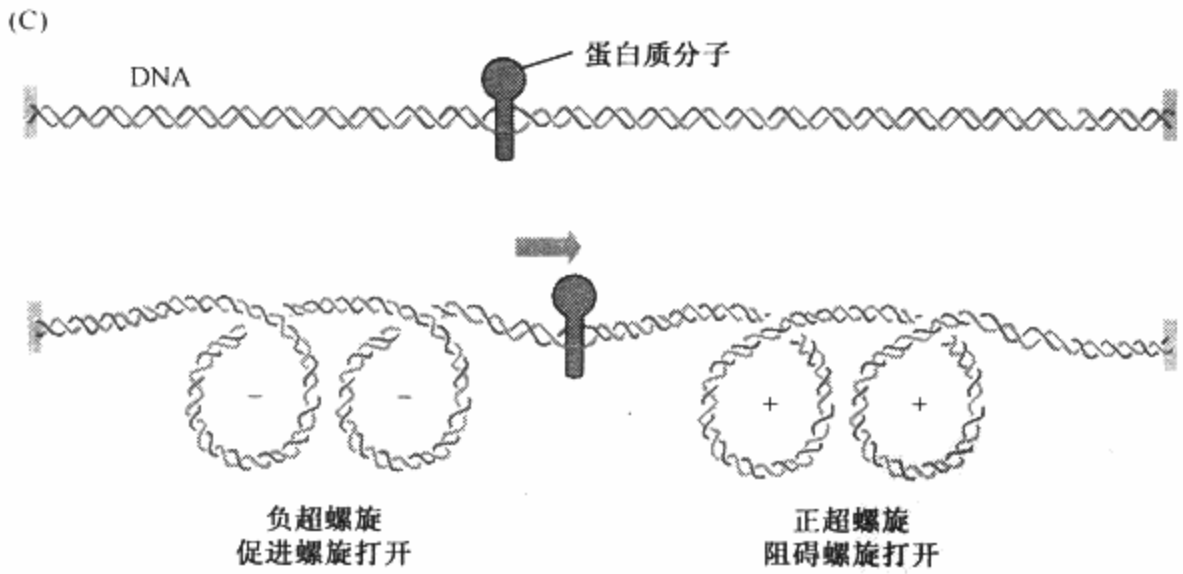


图 6-20　DNA 中的超螺旋张力引起 DNA 发生超螺旋

(A) 对于具有一个自由末端的 DNA 分子而言（或者作为旋轴的一条链上有一个切口），每打开 10 个核苷酸对双螺旋 DNA 就会旋转一圈。(B) 如果旋转受到阻碍，那么就会通过打开螺旋向 DNA 中引入超螺旋张力。一种容纳这种张力的方式是将本例中仍然保持着的双螺旋中的螺旋扭转数从每圈 10 个核苷酸对增加到 11 个；然而，DNA 螺旋抵制这种像弹簧一样的形变，更倾向于通过弯曲成超螺旋环释放超螺旋张力。结果，每打开 10 个核苷酸对，就在 DNA 双螺旋中形成一个 DNA 超螺旋。在这种情况下形成的超螺旋是正超螺旋。(C) 沿 DNA 双螺旋搜寻的蛋白质引入了 DNA 的超螺旋。这里所示的 DNA 两个末端不能相对于彼此自由旋转，并且假定蛋白质分子移动时也阻止其自由旋转。在这些情况下，蛋白质的移动，引起位于蛋白质前面的 DNA 螺旋累积过量的螺旋回转，而蛋白质后面 DNA 中缺乏螺旋回转，如图所示

的螺旋（就好像它们位于一个 DNA 环中，如细菌染色体；或者位于一个牢牢地夹住的环，据认为它存在于真核染色体中），作为补偿，它就会形成一个巨大的 DNA 超螺旋。这个超螺旋的形成在能量上是有利的，因为它将保持碱基配对的区域恢复成正常的螺旋缠绕，否则的话，这个区域由于末端固定了，就需要发生过度盘绕。

当 RNA 聚合酶沿一段 DNA 移动时，由于它的末端锚定，所以也会产生超螺旋张力[图 6-20(C)]。只要聚合酶不能迅速地自由旋转（考虑到 RNA 聚合酶的大小以及它们附着的转录物，这种旋转是不可能发生的），移动的聚合酶就会在它前面的 DNA 中产生正超螺旋张力，而在后面的 DNA 中产生负超螺旋张力。对于真核生物，我们认为这种情况提供了一种意外收获：聚合酶前面的正超螺旋使 DNA 螺旋更难打开，但是当 DNA 从组蛋白核心释放，帮助松弛正超螺旋张力时，这种张力能促进核小体中的 DNA 解开包装。

任何蛋白质独自推动自身沿双螺旋的一条 DNA 链前进，都会产生超螺旋张力。在真核生物中，DNA 拓扑异构酶迅速地将这种超螺旋张力消除。但是在细菌中，一种称为 DNA 旋转酶的特殊拓扑异构酶利用 ATP 水解的能量，不断向 DNA 引入超螺旋，从而使 DNA 保持恒定的张力。引入的这些都是负超螺旋，它们与一段 DNA 螺旋区域打开时形成的正超螺旋的手性相反［图 6-20(B)］。无论什么时候一段螺旋区域打开了，就会从细菌 DNA 消除这些负超螺旋，从而降低超螺旋张力。因此，与未超螺旋化 DNA 中螺旋的打开相比，DNA 旋转酶使得细菌中打开 DNA 螺旋在能量上有利。由于这个原因，它常常促进了细菌中那些需要解开螺旋的遗传过程，包括由细菌 RNA 聚合酶进行的转录起始。

真核生物中转录延伸与 RNA 加工紧密偶联

我们已经看到细菌 mRNA 仅仅由在基因组特定位点起始和停留的 RNA 聚合酶合成。在真核生物中的情况则显著不同。特别是，转录只是一系列反应的第一步。这些反应包括 RNA 两个末端的共价修饰，以及通过 RNA 剪接去除从 RNA 转录物中抛弃的内含子序列（图 6-21）。真核 mRNA 末端的修饰分别是 5′端加帽和 3′端多聚腺苷酸化（图 6-22）。这些特殊末端使细胞在 mRNA 分子输出细胞核用于翻译成蛋白质之前，能评估 mRNA 分子的两个末端是否都存在（因而信息是完整的）。在第 4 章中，我们看到存在于基因组中的典型真核基因，是由很长的内含子分隔开的编码蛋白质的小序列区域。因此 RNA 剪接是至关重要的步骤，在此过程中将不同部分的蛋白质编码序列重新连接起来。下面我们将会讨论到，RNA 剪接也为更高级的真核生物提供了从同一基因合成几种不同蛋白质的能力。

这些 RNA 加工步骤通过一种精巧的机制与转录延伸紧密地偶联起来。前面讨论过，RNA 聚合酶向 RNA 合成延伸模式转变中的一个重要步骤，就是称作 CTD 的 RNA 聚合酶Ⅱ尾巴发生广泛的磷酸化。最大亚基的这个 C 端区域由一长段串联排列的 7 个氨基酸重复序列组成，每个重复含有 2 个丝氨酸，可以磷酸化。由于在人类 RNA 聚合酶Ⅱ的 CTD 中有 52 个重复，所以它完全磷酸化就会在聚合酶上增加 104 个带负电荷的磷酸基团。这个磷酸化步骤不仅使 RNA 聚合酶与转录起始位点上其他蛋白分离，也使一系列新蛋白质能够结合 RNA 聚合酶尾巴，后者在转录延伸和前体 mRNA 加工中都发挥作用。下面将会讨论，这些加工蛋白质中有些似乎在新生 RNA 分子刚从 RNA 聚合酶出来时，就从聚合酶尾巴“跳”到该 RNA 上，开始对它进行加工。因此，可以把处于延伸模式的 RNA 聚合酶Ⅱ看成是一个 RNA 工厂，不仅将 DNA 转录成 RNA，而且加工它合成的 RNA（图 6-23）。

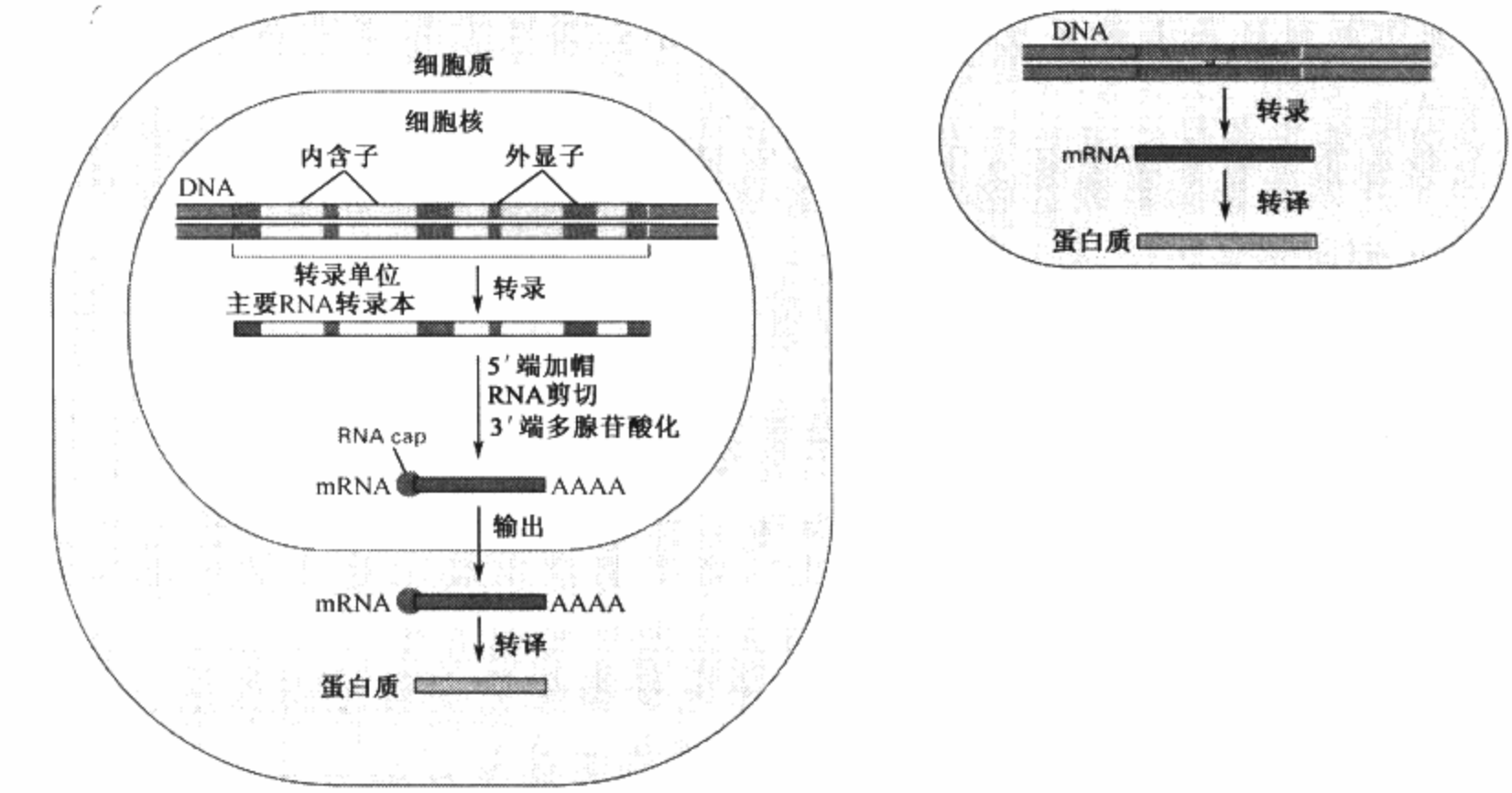


图 6-21　概括真核生物和细菌中从基因到蛋白质经历的步骤

细胞中蛋白质最后的水平决定于每一步的效率和 RNA 与 DNA 分子降解的速率。（A）在真核生物细胞中，只通过转录产生的 RNA 分子（有时指初始转录物）将含有编码（外显子）和非编码（内含子）序列。在它能翻译成蛋白质之前，RNA 两个末端经过修饰，并通过酶催化的 RNA 剪接反应将内含子除去，结果生成的 mRNA 再从核内转运到细胞质中。虽然这些步骤被描述成在逐步地以先后顺序发生，但实际上，它们是偶联起来的，不同的步骤可以同时进行。例如，转录完成之前就加上了 RNA 帽并且剪接通常已开始进行。由于存在这种偶联，所以完整的初始 RNA 转录物通常不存在于细胞内。（B）在原核生物中，mRNA 分子的合成要简单得多。mRNA 分子的 5′端由 RNA 聚合酶起始转录合成，3′端则由转录终止合成。由于原核生物没有细胞核，转录和翻译在共同的空间中进行。实际上细菌 mRNA 的翻译常常在合成结束之前就开始了

RNA 加帽是真核生物前体 mRNA 第一种修饰

RNA 聚合酶一旦合成约 25 个核苷酸的 RNA 之后，就会修饰新 RNA 分子的 5′端，增加一个由修饰后的鸟嘌呤核苷酸组成的帽子[图6-22(B)]。加帽反应由三种酶连续作用来完成：一个（磷酸酶）将新生 RNA5′端的磷酸去掉，另一个（鸟苷转移酶）将一个 GMP 反向连接上来（5′→5′，而不是 5′→3′），第三个（甲基转移酶）在鸟嘌呤上增加一个甲基（图 6-24）。由于所有三种酶都结合到磷酸化的 RNA 聚合酶尾巴上，所以它们准备好一旦新生的转录物从聚合酶上出来就去修饰它的 5′端。

这个 5′甲基帽子提供了真核 mRNA 5′端的信号，此标记有助于细胞区别 mRNA 和细胞内存在的其他类型 RNA 分子。例如，RNA 聚合酶Ⅰ和Ⅲ转录时产生的是未加帽子的 RNA，部分是因为这些聚合酶缺少尾巴。我们将在下面一部分讨论到，在细胞核内这个帽子结合上一个称为 CBC 的蛋白质复合体（帽子结合复合物），由它帮助对 RNA 进行正确的加工并输出细胞核。5′-甲基帽子也在细胞质内的 mRNA 翻译过程中发挥重要作用，这些将在本章后面讨论。

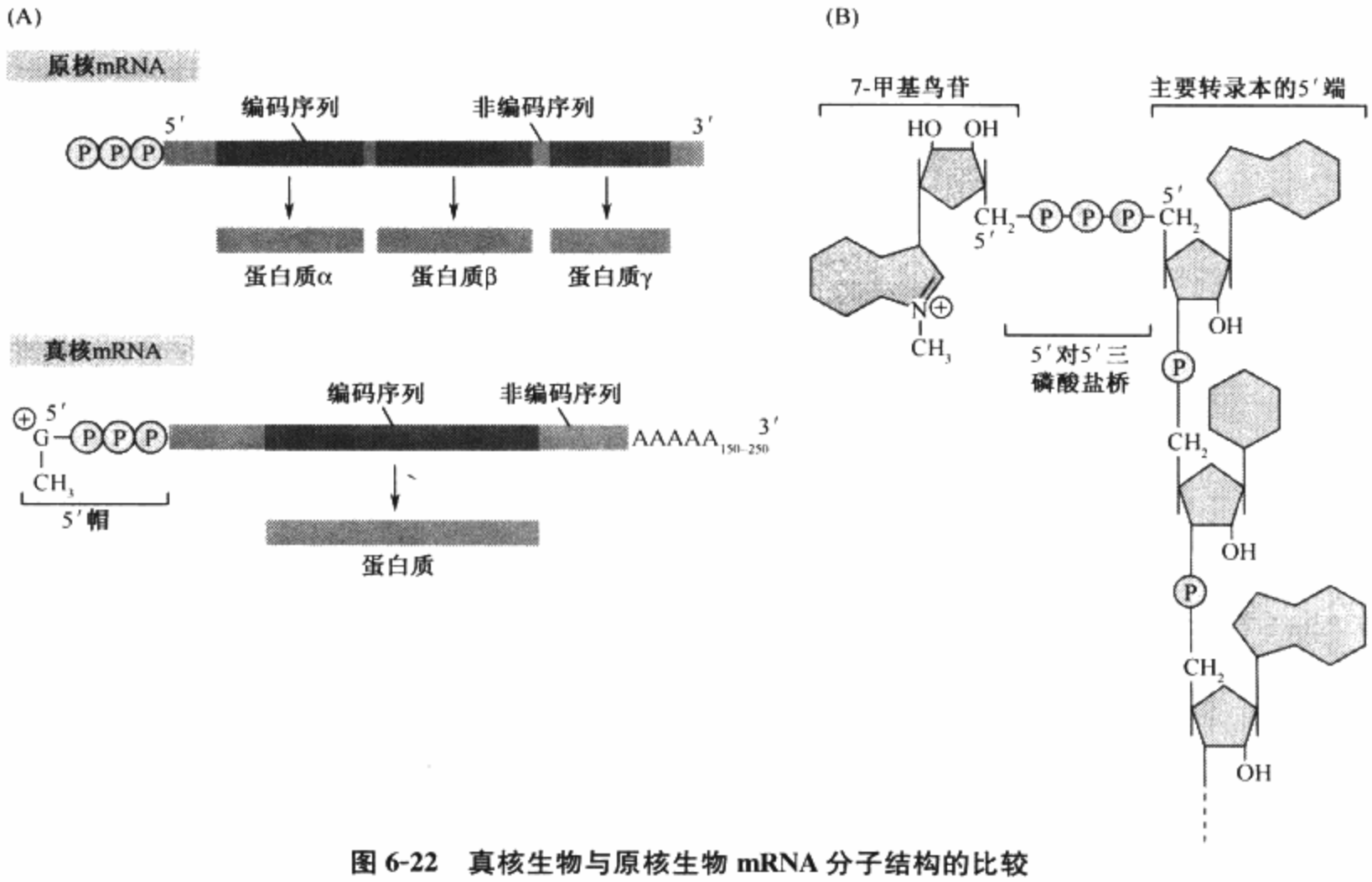


图 6-22　真核生物与原核生物 mRNA 分子结构的比较

（A）细菌 mRNA 5′端和 3′端是 RNA 聚合酶合成的链上未加修饰的末端。该酶在这些位点分别起始和终止转录。真核 mRNA 对应末端的形成是分别通过增加一个 5′帽子，对前体 mRNA 进行切割，以及增加一个多聚 A 尾。该图还说明了真核与原核 mRNA 其他不同之处：细菌 mRNA 能包含许多不同蛋白质的指令，而真核 mRNA 几乎总是只含有一种蛋白质的指令。（B）真核 mRNA 分子 5′端帽子的结构。注意 7-甲基 G 与 RNA 末端之间的异常的 5′→5′连接。许多真核 mRNA 携带有另一种修饰：mRNA 中第二个核糖 2′—OH 基团发生甲基化（未示出）

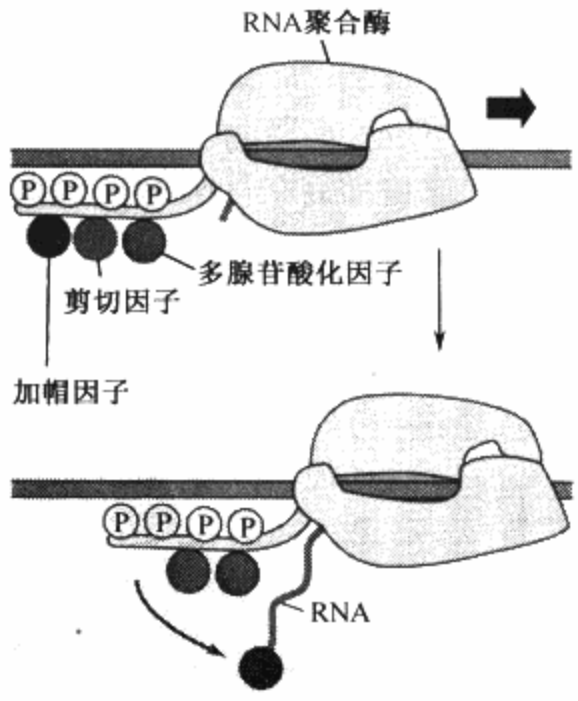


图 6-23　真核 RNA 聚合酶Ⅱ的“RNA 工厂”观念

RNA 聚合酶Ⅱ不仅将 DNA 转录成 RNA，而且在其尾部还携带着前体 mRNA 加工蛋白质，它们在适当的时候转移到新生的 RNA 上。存在许多种 RNA 加工酶，但并不是所有的都随聚合酶一起移动。对于 RNA 剪接，如尾部只携带有少数关键的组分；一旦转移到 RNA 分子上，它们就作为成核位点，供其余组分结合。在转录起始过程后期，当 RNA 聚合酶尾部发生磷酸化时，RNA 加工蛋白质首先与它结合（图 6-16）。一旦 RNA 聚合酶Ⅱ完成转录，就从 DNA 上释放下来，其尾部磷酸由可溶的磷酸酶除去，这样它就能再次起始转录。只有这种去磷酸化的 RNA 聚合酶Ⅱ形式才能胜任在启动子上开始 RNA 合成

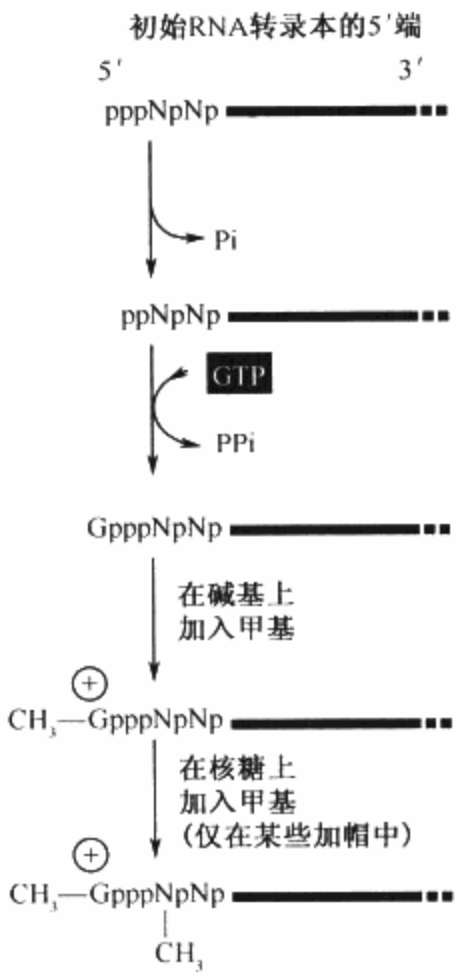


图 6-24　将 RNA 聚合酶Ⅱ合成的每个 RNA 分子 5′端加帽的反应

最终的帽子含有一个少见的 5′→5′连接，位于带正电荷的 7-甲基 G 残基和 RNA 转录物 5′端之间[图 6-22(B)]。字母 N 代表 4 种核苷酸中任一种，尽管开始一条 RNA 链的核苷酸常常是一个嘌呤(一个 A 或 G)(摘自 A. J. ShatKin, *BioEssays* 7:275～277,1987. © ICSU Press)

RNA 剪接从新转录的前体 mRNA 中除去内含子序列

在第 4 章讨论过，真核基因的蛋白质编码序列通常由间隔的非编码序列（内含子）打断。1977 年发现真核基因的这个特征，对于科学家来说是一个出乎意料的事情，因为直到那时，他们一直熟知细菌的基因通常由一段连续的编码 DNA 组成，这些 DNA 直接转录成 mRNA。形成鲜明对比的是，科学家发现真核基因都打断成小片段的编码序列（表达序列或**外显子**），与长得多的干扰序列或**内含子**间隔存在；因此真核基因的编码序列常常只占整个基因长度的一小部分（图 6-25）。

无论内含子和外显子都转录成 RNA。内含子序列通过**RNA 剪接**过程从新合成的 mRNA 中除去。细胞内进行的大多数 RNA 剪接都在合成 mRNA 过程中发挥作用，我们有关剪接的讨论也主要集中在这种类型上。将它称作前体 mRNA（或 pre-mRNA）剪接是为了表示它在那些注定成为 mRNA 的 RNA 分子上进行。只有在 5′端和 3′端加工和剪接完成之后，这种 RNA 才称为 mRNA。

每次剪接除去一个内含子，此过程通过连续的磷酸转移反应，即转酯反应进行；这些转酯反应将两个外显子连接起来，同时以“套索”形式除去内含子（图 6-26）。由于磷酸键数目保持未变，从理论上讲这些反应不需要核苷三磷酸水解也能发生。然而，催化前体 mRNA 剪接的装置是很复杂的，它由 5 种其他 RNA 分子以及超过 50 种蛋白质组成，而且每次剪接它都水解许多 ATP 分子。可能是需要这种复杂性来确定剪接是高度精确的，而同时也足够灵活，以应对典型真核细胞内出现的大量不同内含子。RNA 剪接频繁出现差错将严重伤害细胞，因为这些错误将导致蛋白质功能失常。第 7 章我们将看到，当罕见的剪接错误确实出现以后，细胞有一个“纠错”装置用于清除未正确剪接的 mRNA。

由 RNA 剪接除去大量内含子看起来似乎很浪费。为了努力解释它为什么会发生，科学家已指出，外显子-内含子这种排列似乎促进了新的有用蛋白质的出现。因此 DNA 中存

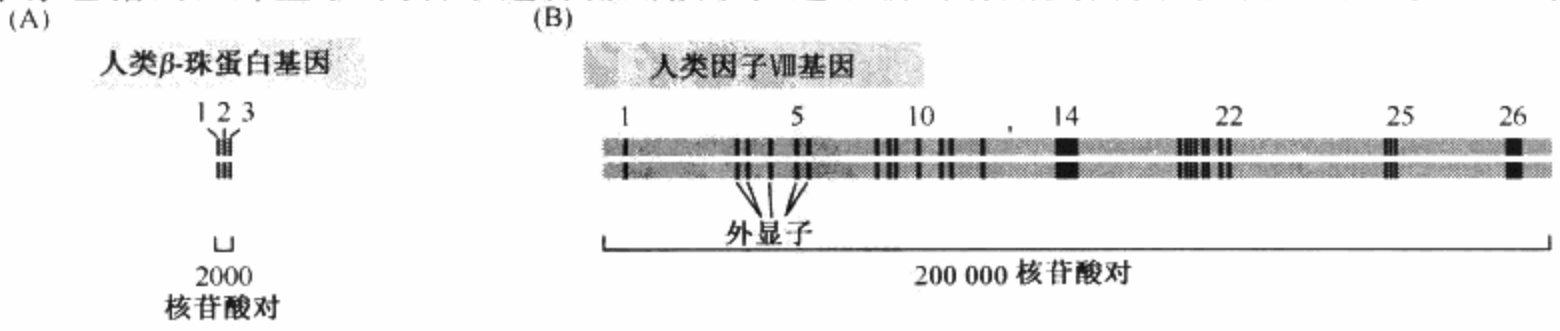


图 6-25　显示外显子和内含子排列的两个真核基因结构

（A）比较小的编码输氧血红蛋白一个亚基的 β-珠蛋白基因含有 3 个外显子（图 4-7）。（B）大得多的因子Ⅷ基因含有 26 个外显子；它编码一个在血凝途径中起作用的蛋白质（因子Ⅷ）。该基因的突变导致最普遍的血友病形式

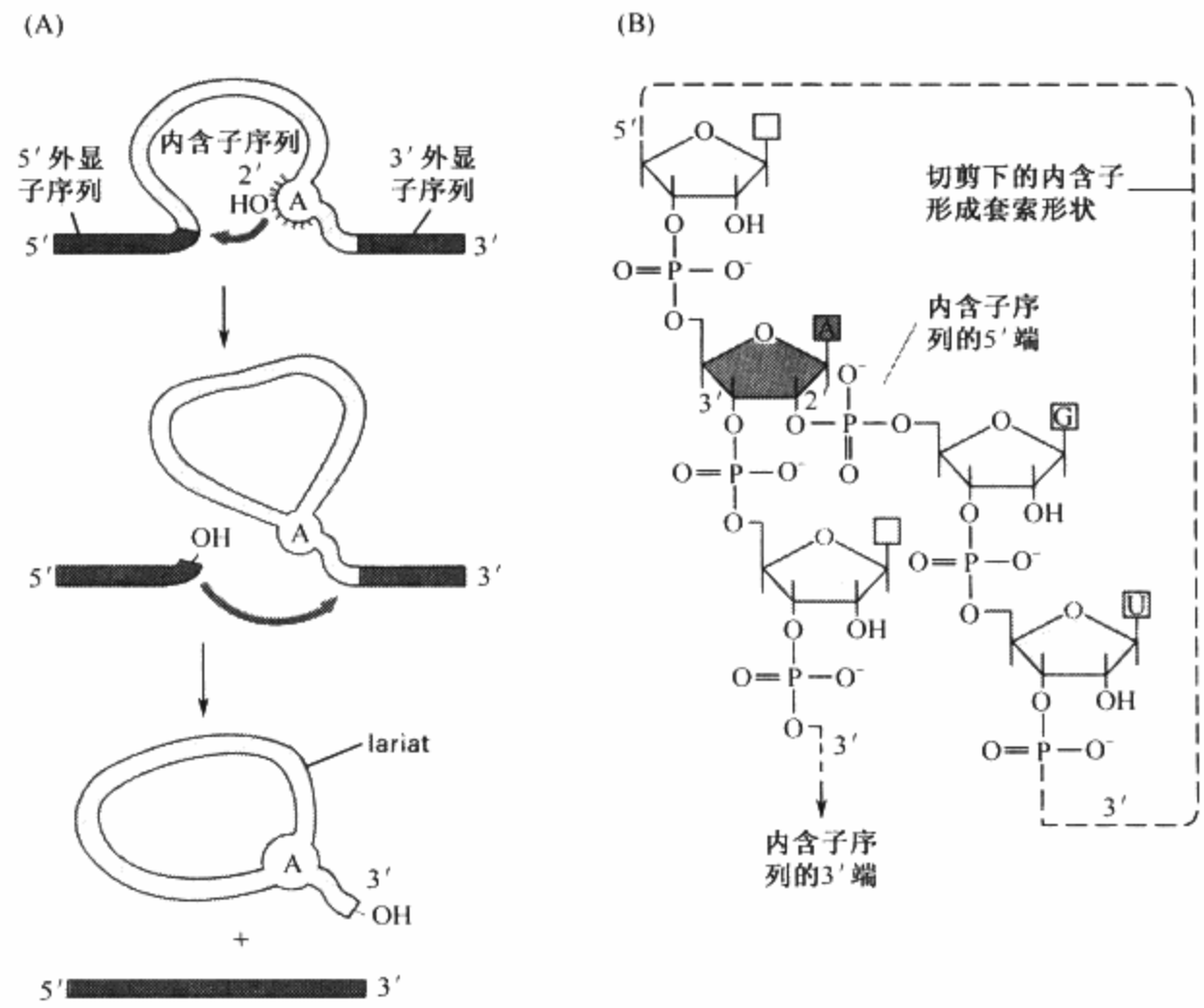


图 6-26　RNA 剪接反应

（A）第一步，内含子序列中特定的一个腺嘌呤核苷酸（用红色表示）攻击 5′剪接位点，并在此位点将 RNA 的糖-磷酸骨架切开。切开的内含子 5′端与腺嘌呤核苷酸共价连接，如（B）中详细所示，因而在 RNA 分子中产生一个环。释放的外显子自由 3′—OH 端然后与下一个外显子起始处发生反应，从而将两个外显子连接起来，并以套索形状释放出内含子序列。因而两个外显子序列连接成一个连续的编码序列，释放的内含子序列则在适当的时候被降解掉

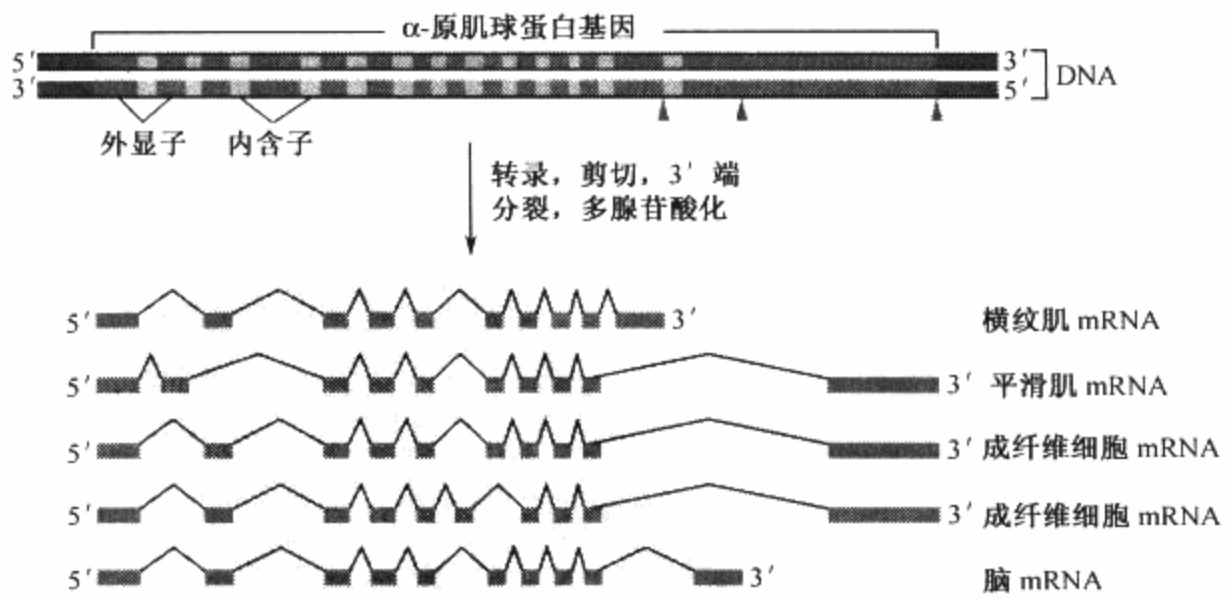


图 6-27　大鼠 α-原肌球蛋白基因的选择性剪接

α-原肌球蛋白是在肌肉细胞中调节收缩的一种卷曲螺旋蛋白质。初始转录物可以不同的方式剪接，如图所示，产生不同的 mRNA，然后由它们生成不同的蛋白质。有些剪接模式是某些细胞类型特异的。例如，横纹肌中合成的 α-原肌球蛋白与平滑肌中由相同基因合成的 α-原肌球蛋白不同。图上部的箭头表示发生切割和加多聚 A 的位点

在的大量内含子使基因重组能够容易地将不同基因的外显子组合起来。从而通过现存基因部分的组合，使新的蛋白质编码基因更容易进化出来。第 3 章已有描述，现今的许多细胞就像是由一套称作蛋白质结构域的共有蛋白质片段组成的拼装物，这一现象支持了上述观点。

RNA 剪接具有一个现时优势。许多真核基因转录物通过多种方式进行剪接，产生一套不同的 mRNA，因而能够从同一基因合成一套相应的不同蛋白质（图 6-27）。第 7 章我们讨论选择性剪接的其他例子，因为这也是细胞用来改变它们基因表达的机制之一。RNA 剪接不仅不像乍看起来那样，是一个浪费的过程，它反而使真核细胞能够增加基因组已经很巨大的编码潜力。本章以及下一章，我们将多次回到这个问题上来，但是首先我们需要描述完成这项重大任务的细胞机制。

核苷酸序列为在哪里进行剪接提供了信号

内含子大小在 10 个核苷酸到 100 000 多个核苷酸之间不等。面对一个完整的真核生物基因组，科学家很难找出内含子精确的序列边界（即便有电脑的帮助）。选择性剪接的可能性使得仅从一个基因组序列推测蛋白质序列的难题更加复杂。这种困难构成了确定一个完整基因组序列中全部基因的主要障碍。它也是为什么我们只知道人类基因组，如基因数目的近似值的原因。可是我们体内的细胞却能迅速地以高度忠实性识别和切除适当的内含子序列。我们已经看到了内含子序列的去除涉及 RNA 上 3 个位点：5′ 剪接位点、3′剪接位点和形成切除的套索基本结构的内含子序列中的分支位点。在前体 mRNA 剪接过程中，这 3 个位点每个都有在内含子之间相似的共有序列，为细胞在哪里进行剪接提供线索（图 6-28）。然而，每个序列中也存在足够的变化，以至于让科学家们难以在基因组序列中找出所有的剪接信号。

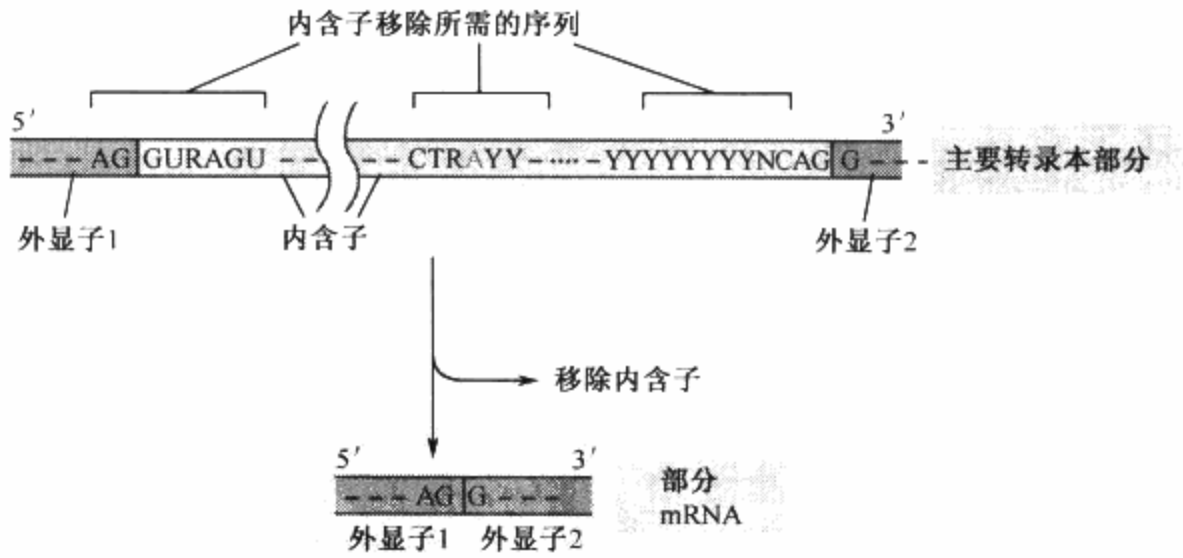


图 6-28　RNA 分子中代表人类大多数内含子起点和终点信号的共有核苷酸序列

图中所示只有三段核苷酸序列是除去内含子序列所必需的；内含子其余部分可以是任何核苷酸。这里 A、G、U 和 C 是标准的 RNA 核苷酸；R 代表 A 或 G；Y 代表 C 或 U。用红色突出的 A 形成剪接产生的套索分支点。只有位于内含子起点的 GU 和末端的 AG 是剪接共有序列中不变的核苷酸。剩余位点（甚至是分支位点 A）可以是不同的核苷酸，尽管图中所示的核苷酸优先存在。三个剪接共有序列之间在 RNA 上的间隔距离高度变化；然而，分支点和 3′剪接点之间的距离通常比 5′剪接点和分支点之间的距离要近得多

RNA 剪接由剪接体完成

与我们已经讨论过的 mRNA 合成中其他步骤不同，RNA 剪接主要由 RNA 分子而不是蛋白质来完成。RNA 分子识别内含子-外显子边界，并参与剪接化学反应。这些 RNA 分子比较短（每个不到 200 个核苷酸），且总共有 5 种（U1、U2、U4、U5 和 U6）参与了前体 mRNA 主要的剪接形式。它们称被为 **snRNA（核内小 RNA）**，其中每个都与至少 7 种蛋白质亚基形成复合物，称作 snRNP（核内小核糖核蛋白）。这些 snRNP 组成了**剪接体**的核心，后者是由细胞内参与前体 mRNA 剪接的 RNA 和蛋白质形成的巨大组装物。

剪接体是一个动态装置：下面我们会看到，它由独立的组分在前体 mRNA 上组装起来，这些组分进入后，随着剪接反应的进行，会再次离开它（图 6-29）。在剪接反应中，对 5′剪接位点、分支位点、3′剪接位点的识别主要通过 snRNA 与前体 mRNA 底物的共有 RNA 序列之间碱基配对完成（图 6-30）。在剪接过程中，剪接体经历几次移位，其间一个碱基配对相互作用断裂，然后在此位置处再形成另外一个碱基对。例如，在 5′剪接点 U6 取代了 U1[图6-30(A)]。我们将看到，在剪接反应中，多次发生这种 RNA-RNA 重排（其间形成 RNA-RNA 相互作用，需要破坏另一个 RNA-RNA 相互作用）。这种重排允许在化学反应能够进行之前，对 RNA 序列进行核查和再次核查，因而提高了剪接的精确性。

剪接体利用 ATP 水解提供能量形成各种复杂的 RNA-RNA 重排

虽然本质上 RNA 剪接不需要 ATP 水解，但是剪接体的逐步组装和重排却需要水解 ATP。其他组成核小体的蛋白质中有些是 RNA 解螺旋酶，它们利用 ATP 水解释放的能量破坏已存在的 RNA-RNA 相互作用，以便形成新的相互作用。实际上，图 6-29 所示的所有步骤（除 BBP 结合到分支位点以及 U1 snRNP 结合到 5′剪接位点之外）都需要 ATP 水解和其他的蛋白质。总的来说，每次剪接需要超过 50 种蛋白质，包括那些形成 snRNP 的蛋白质。

剪接体中进行需要 ATP 的 RNA-RNA 重排，发生在 snRNP 内部和 snRNP 与 mRNA 底物之间。这些重排的重要作用之一就是产生剪接体的活性催化位点。剪接体组分只有在前体 mRNA 底物上完成组装和重排之后才产生活性位点，这种策略是防止产生任意剪接的重要方式。

剪接体最让人惊奇的特征可能就是催化位点本身的实质。它大部分（如果不是绝对的话）由 RNA 而不是蛋白质形成。在本章最后一部分我们从整体上讨论赋予 RNA 催化功能的结构和化学特性。这里我们只需要考虑 U2 snRNA 和 U6 snRNA 在剪接体中形成了一个精确的三维 RNA 结构，将前体 mRNA 5′剪接位点和分支位点并列起来，并且可能发生了第一个转酯反应[见图6-30(C)]。5′和 3′剪接点也以同样的方式并到一起（该过程需要 U5 snRNA）以利于第二个转酯反应的进行。

一旦完成了剪接化学反应，snRNP 仍然结合在套索上，剪接产物释放。这些 snRNP 从套索上（以及彼此之间）解离需要一系列其他需要水解 ATP 的 RNA-RNA 重排，从而使 snRNP 回到原先的构型，以便在新的反应中再次得到利用。

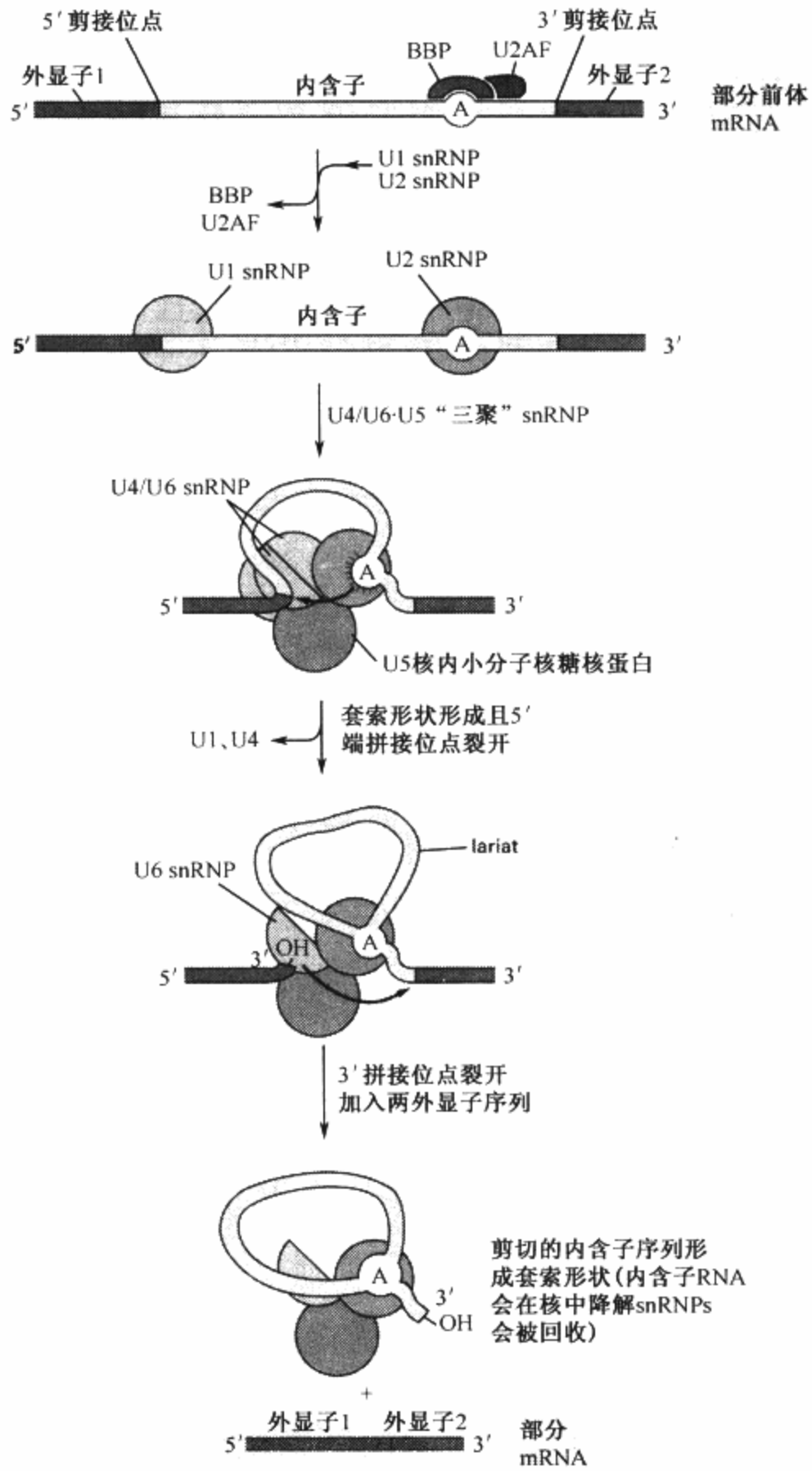


图 6-29　RNA 剪接机理

RNA 剪接由 snRNP（彩色环所示）的组装物和其他蛋白质（大多数未示出）催化，它们共同形成剪接体。剪接体识别前体 mRNA 分子上的剪接信号，将内含子的两个末端并到一起，并提供酶活性催化两步反应（图 6-26）。首先 BBP（分支点结合蛋白）和 U2AF（一种辅助蛋白质）识别分支位点。接着，U2 snRNP 取代 BBP 和 U2AF 并与分支位点共有序列形成碱基配对，而 U1 snRNP 与 5′剪接位点形成碱基配对（图 6-30）。此时，U4/U6/U5 三联体 snRNP 进入剪接体。在此三联体 snRNP 中，碱基配对相互作用将 U4 snRNP 和 U6 snRNP 牢固地维系在一起，U5 snRNP 的结合则更加松弛一些。然后发生多次 RNA-RNA 重排，将 U4/U6 碱基配对打开（如图所示，U4 snRNP 在剪接完成之前从剪接体排除），并使 U6 snRNP 能在 5′剪接点取代 U1（图 6-30）。随后进行的重排产生剪接体的活性位点，并将前体 mRNA 底物适当部分定位以便进行剪接反应。尽管在图中未示出，但每个剪接事件还需要其他的蛋白质，它们有些水解 ATP 并促进 RNA-RNA 重排

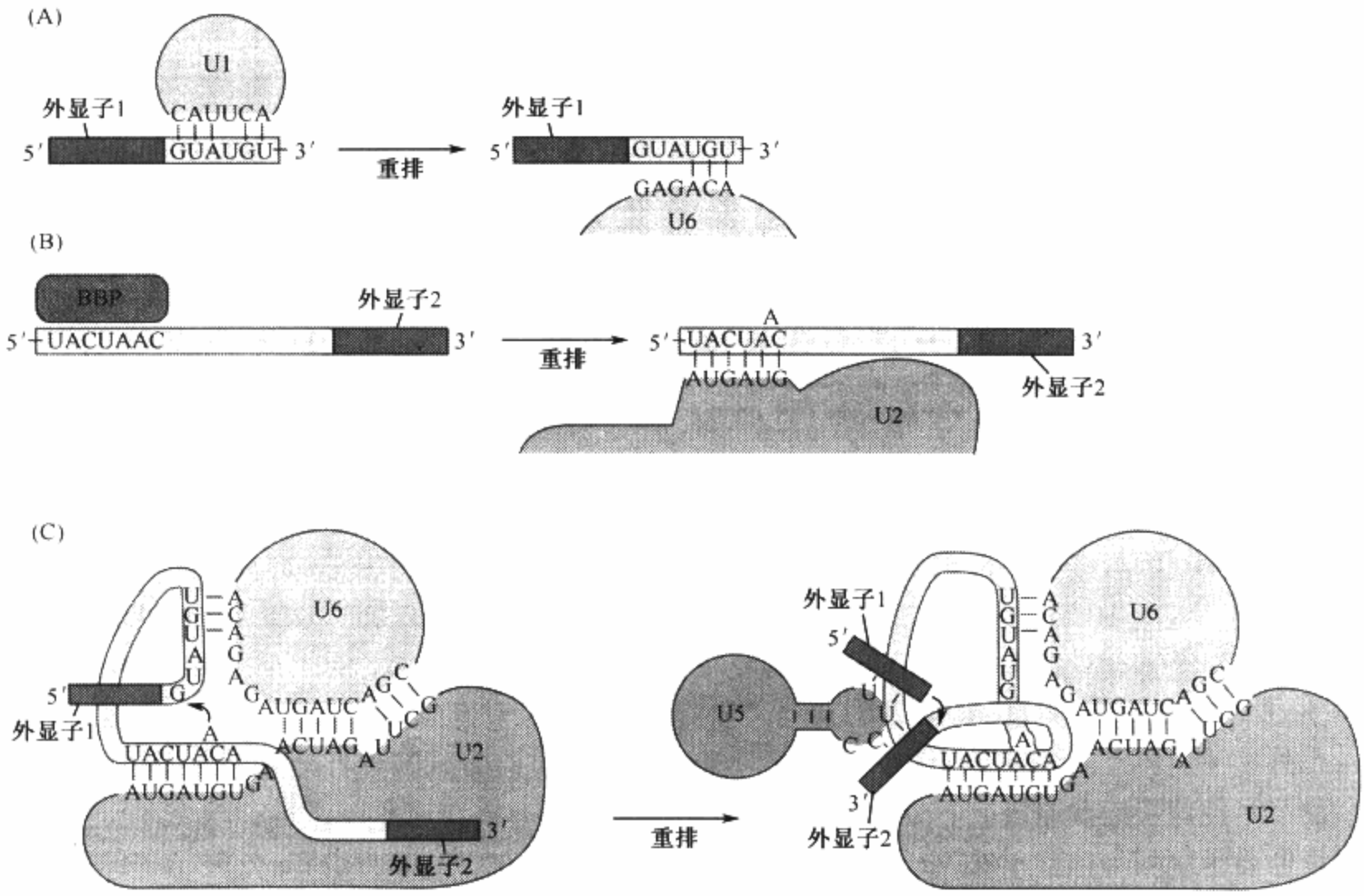


图 6-30　前体 mRNA 剪接过程中剪接体内发生的多次重排

这里显示的是酿酒酵母中的详细情况，其中涉及的核苷酸序列与人类细胞中的稍有不同。（A）U6 snRNP 交换 U1 snRNP 在第一个磷酸转移反应之前进行（图 6-29）。这个交换使 5′剪接位点能由两种不同的 snRNP 阅读，因而增加了剪接体选择 5′剪接位点的精确性。（B）分支位点首先由 BBP 识别，然后再由 U2 snRNP 识别；与（A）中一样，这个“检验和再检验”策略增加了位点选择的精确性。U2 对分支点的结合迫使适当的腺嘌呤核苷酸（红色）不能配对，因而激活它去攻击 5′剪接位点（图 6-29）。这与 BBP 的识别一起，成为剪接体精确选择最终形成分支位点的腺嘌呤核苷酸的方式。（C）第一次磷酯转移反应（左边）之后，U5 snRNP 经历重排，将两个外显子带至相邻位置，以便进行第二次磷酯转移反应（右边）。snRNP 即将反应物定位，又提供两个反应的催化位点（全部或者部分）。U5 snRNP 在发生重排以前就存在于剪接体；为了清楚起见，已从左图中略去。正如文中所讨论的那样，图中显示的所有 RNA-RNA 重排（以及其他在剪接体中发生但未示出的）需要其他蛋白质和 ATP 水解作用的参与

对前体 mRNA 的有序影响有助于解释如何选择正确的剪接位点

我们已经看到，内含子序列大小变化非常大，有些超过了 100 000 个核苷酸。如果剪接位点的选择只由作用在预先形成的无结合蛋白质 RNA 分子上的 snRNP 决定，那么可想而知，剪接错误（如外显子跳越以及利用隐蔽剪接位点）将会频繁发生（图 6-31）。

剪接体固有的忠实机制得到其他两种因素的补充，它们有助于确保剪接精确地进行。这些对前体 RNA 的有序影响，增加了在拼接体内发生剪接化学反应之前，将一对合适的 5′和 3′ 剪接位点并到一起的可能性。第一因素来源于前体 mRNA 从正在转录的 RNA 聚合酶 Ⅱ（图 6-23）出现时，发生的剪接体组装。与形成 5′帽子的情况一样，似乎多个剪接体组分都携带在 RNA 聚合酶的磷酸化尾巴上。它们直接从聚合酶向新生前体 mRNA 的转移可能有助于细胞掌握内含子和外显子的动态情况：起初 5′剪接位点的 snRNP 只与一个 3′剪接位点共同存在，因为下游的位点还没有合成出来。这个特征有利于防止不正确的外显子跳越。

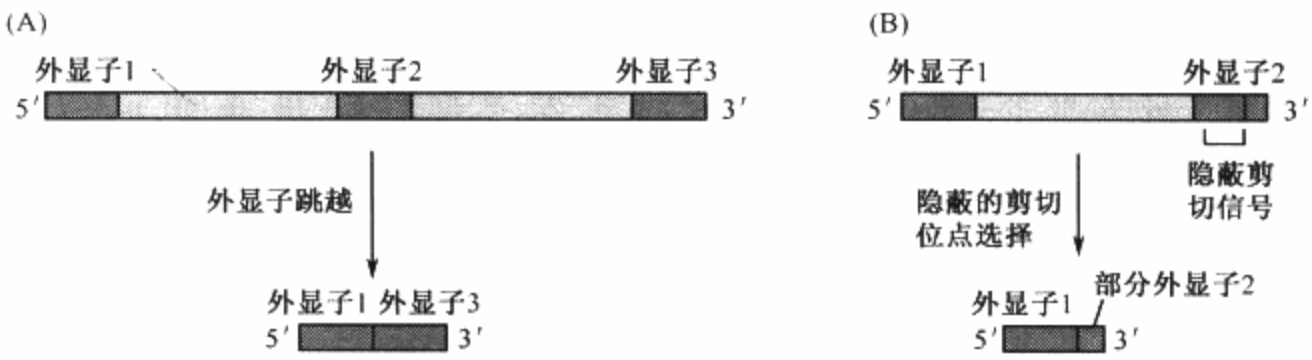


图 6-31　两类剪接错误

如果剪接体在一个预先形成的，无蛋白质的 RNA 分子上进行剪接位点选择，预计这两类错误都可能频繁发生。“隐蔽”剪接信号是与真正的剪接信号特别类似的 RNA 核苷酸序列

第二个帮助细胞选择剪接位点的因素称为“外显子定义假说”，对它还只有大致的了解。外显子大小比内含子大小似乎更加一致，在各种真核生物中的平均大小约为 150 个核苷酸对（图 6-32）。随着 RNA 合成的进行，据认为有一组剪接体成分，称为 SR 蛋白（这样命名是因为它们有一个富含丝氨酸和精氨酸的结构域），在外显子序列上组装起来，从 RNA 5′端开始，标出每个 5′和 3′剪接位点（图 6-33）。这种组装与 U1 snRNA 和 U2 AF 协力进行，前者标记一个外显子边界，后者帮助标记另一个外显子边界。通过这方式特异性地标记外显子，细胞增加了早期剪接组分组装到新生 RNA 上的精确性，因而有助于避免隐蔽剪接位点。SR 如何区分外显子序列与内含子序列还不清楚，然而有些 SR 蛋白优先结合到特定外显子中的 RNA 序列上。从理论上讲，在进化过程中可以利用遗传密码的冗余，为 SR 蛋白选择出外显子中的结合位点，使得产生这些位点的同时，不会限制氨基酸序列。

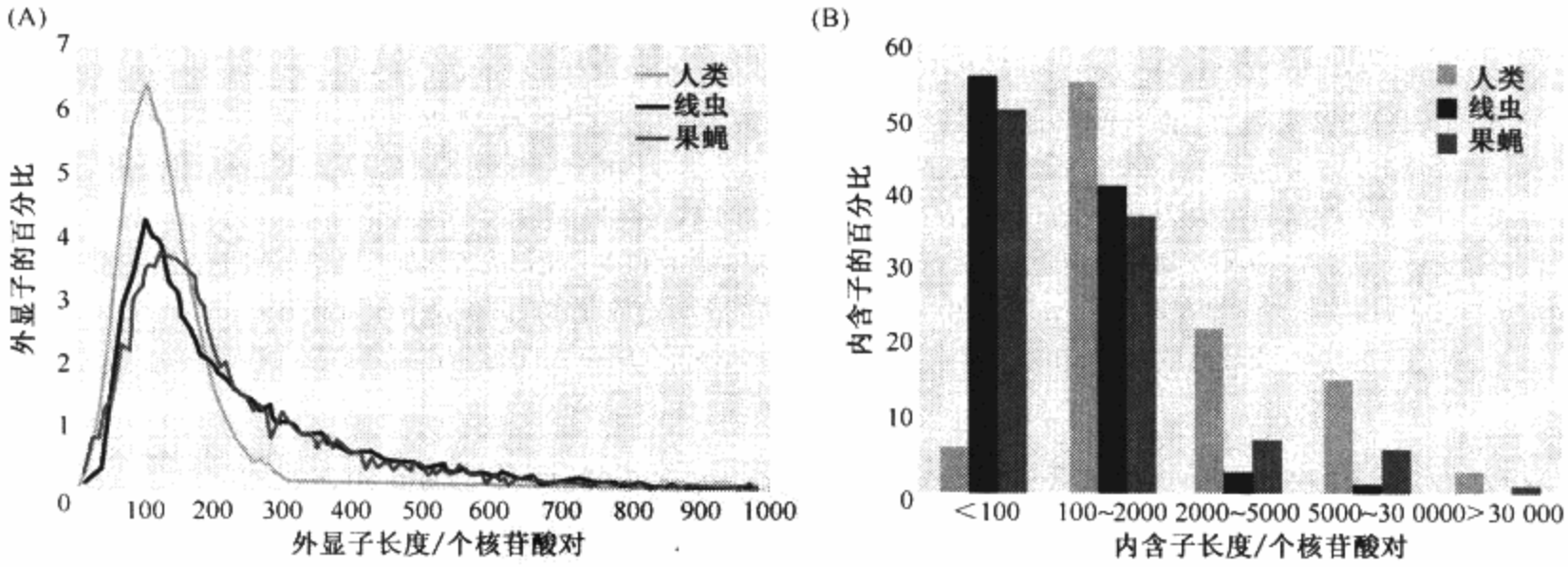


图 6-32　人类、线虫和果蝇基因组中内含子和外显子的长度变化

（A）外显子大小分布。（B）内含子大小分布。注意外显子长度比内含子长度要一致得多（摘自 International Human Genome Sequencing Consortium，*Nature* 409：860～921，2001）

无论是外显子和内含子边界的标记，还是剪接体的组装，都在 RNA 3′端还正被 RNA 聚合酶延伸时就在它上面开始了。然而实际上剪接化学反应的发生可以晚很多。这种延迟意味着内含子序列并不一定要按它们沿 RNA 链排列的先后顺序从前体 mRNA 分子中去除。它也意味着，尽管剪接体组装与转录同时进行，可是有时候剪接反应是在转录后进行的，也就是说，在一个完整的前体 mRNA 分子合成出来以后。

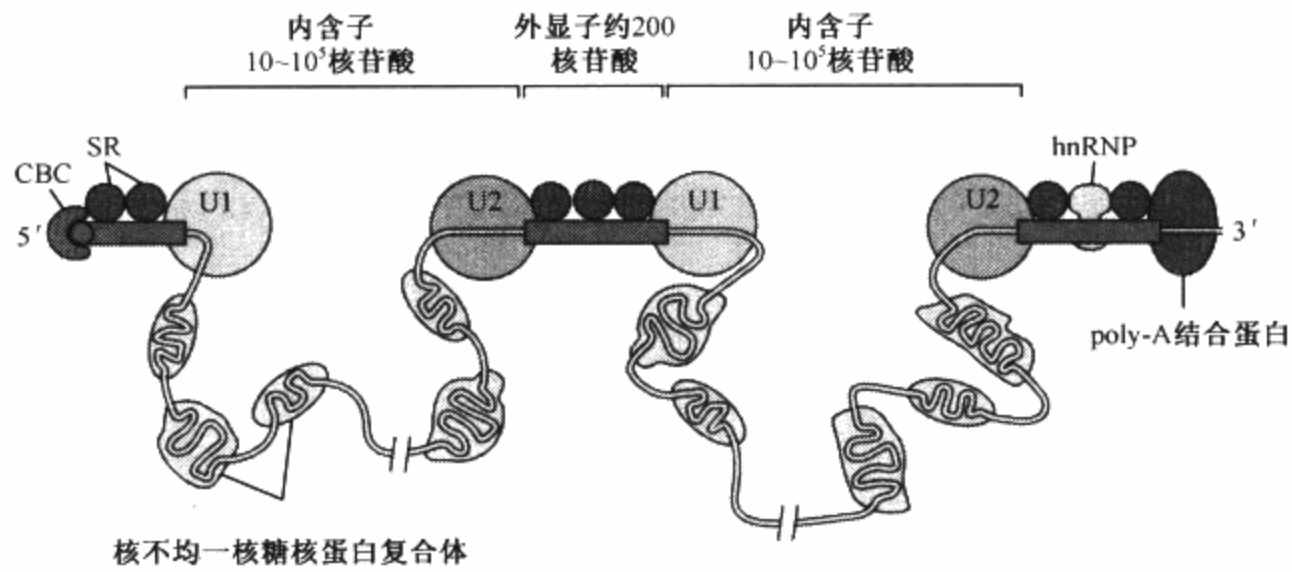


图 6-33　外显子界定假说

根据一种提议，SR 蛋白结合前体 mRNA 中的每个外显子序列，因而有助于指导 snRNP 到达正确的内含子/外显子边界。SR 蛋白划定外显子界线是共转录的，开始于 5′端的 CBC（帽子结合复合物）。如图所示，前体 mRNA 中的内含子序列可以很长，它们包装成 hnRNP（核内不均一核糖核蛋白）复合物，这些复合物将它们压缩成更加有序的结构并可能掩盖了隐蔽剪接位点。每个 hnRNP 复合物形成一个改为核小体直径 2 倍的颗粒，其核心由至少 8 个不同的蛋白质组成。有人提出 hnRNP 蛋白优先与内含子序列结合，这种优先性也有助于剪接体识别内含子和外显子。然而，如图所示，至少有些 hnRNP 蛋白可能结合外显子序列，但是它们在外显子界定中的作用，如果有的话，还有待确立（改编自 R. Reed，*Curr*，*opin. Cell. Biol.* 12：340～345，2000）

另外一套 snRNP 剪接了动物和植物中一小部分内含子

简单的真核生物，如酵母，只有一套 snRNP，并由它们完成全部前体 mRNA 剪接。然而，更复杂的真核生物，如果蝇、哺乳动物和植物拥有另外一套 snRNP，它们指导一小部分内含子的剪接。这种次要的剪接体形式识别 5′和 3′剪接点和分支位点上不同的一套 DNA 序列；由于它的内含子-外显子边界的核苷酸序列决定子，所以称它为 AT-AC 剪接体（图 6-34）。虽然识别不同的核苷酸序列，但是在这种剪接体中的 snRNP 和前体 mRNA 之间以及彼此之间产生的 RNA-RNA 相互作用，与主要剪接体中产生的相互作用是一样的[图6-34(B)]。最新发现的这类 snRNP 使我们对有关主要剪接体中产生的碱基配对相互作用的推理更加肯定，因为虽然涉及的 RNA 序列不同，但是它提供了一套独立的产生相同 RNA-RNA 相互作用的分子。

在少数真核生物中，发现了剪接的一种特殊变化形式，称为**反式剪接**。这些生物包括单细胞的椎虫（在人类中引起非洲睡眠疾病的原生动物），以及多细胞模式生物——线虫。在反式剪接中，来自两个单独的 RNA 转录物的外显子剪接在一起，形成一条成熟的 mRNA 分子（图 6-34）。椎虫的所有基因都以这种方式产生，而线虫 mRNA 只有大约 1%是通过反式剪接产生的。在这两个例子中，一个外显子被剪接到细胞产生的许多不同 RNA 转录物的 5′端；通过这种方式，所有的反式剪接产物都有相同的 5′外显子和不同的 3′外显子。许多在常规剪接过程中发挥作用的 snRNP 也在此反应中得到利用，显然反式剪接利用了携带在共同外显子中的独特 snRNP（称为 SLRNP）（图 6-34）。

少数生物利用反式剪接的原因还不清楚；但是有人认为 5′端共同的外显子可能有助于 mRNA 的翻译。因此，在线虫中反式剪接产物的翻译效率似乎特别高。

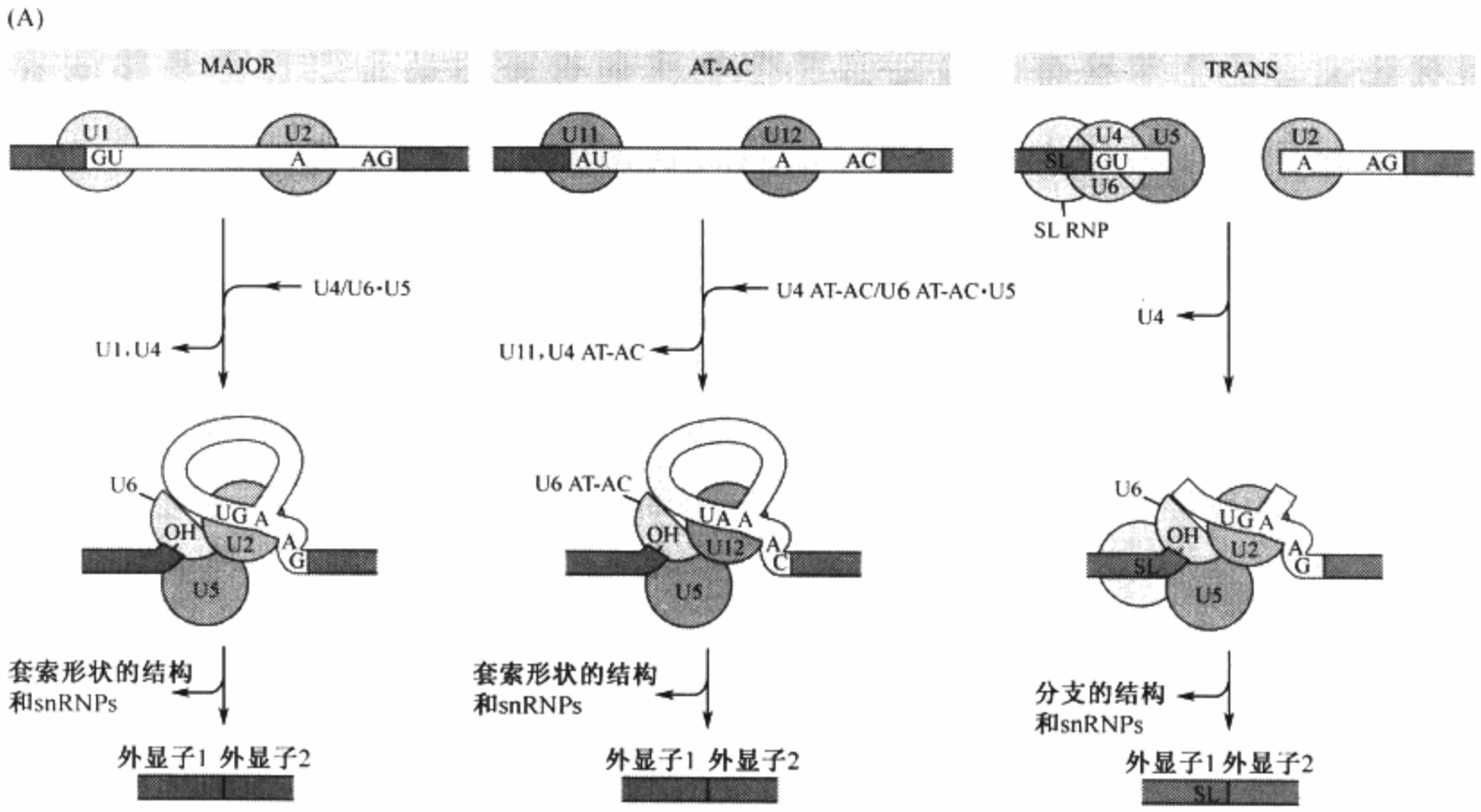


图 6-34　三种 RNA 剪接机制简图

（A）三种剪接体。显示了各自处于两个组装阶段的主要剪接体（左边）、AT-AC 剪接体（中间）和反式剪接体（右边）。U5 snRNP 是这三种剪接体共有的唯一组分。由 AT-AC 剪接体除去的内含子与主要剪接体除去的内含子有不同的共有核苷酸序列。在人类中，估计 0.1%的内含子由 AT-AC 剪接体除去。在反式剪接过程中，SL snRNP 在反应中消耗掉，因为 SL snRNP 一部分成为成熟 mRNA 的第一个外显子。（B）主要的 U6 snRNP 和 U6 AT-AC snRNP 都识别 5′剪接点，但它们通过一套不同的碱基配对相互作用来完成。这里所示的序列来自人类（摘自 Y. -T. Yu et al., The RNA World, pp. 487～524. Cold Spring Harbor, New York: Cold Spring Harbor laboratory Press, 1999）

RNA 剪接表现出显著的可塑性

我们已经看到剪接位点的选择取决于前体 mRNA 转录物的许多特征，包括 RNA 上面的三种信号对剪接装置的亲和力（5′和 3′剪接点以及分支位点）、外显子的长度和核苷酸序列、剪接体的共转录组装、界定外显子的“规则”的精确性。到目前为止，我们已经强调了在细胞内进行的 RNA 剪接过程的精确性。但是这种机制的选择似乎也因为它的柔性，它使细胞偶尔能够试验合成新的蛋白质。因此，比如说，对特定的内含子剪接很关键的核苷酸序列发生突变时，并不一定完全阻止这个内含子的剪接。相反，突变通常产生一种新的剪接模式（图 6-35）。更常见的是，仅仅跳过了外显子[图6-35(B)]。在其他情况下，突变导致利用一个“隐蔽”剪接位点[图6-35(C)]。可能经过进化，剪接装置已能挑选出最可能的剪接衔接点模式，如果由于突变破坏了最佳的选择，它就会选择下一个最好的，以此类推。在 RNA 剪接过程中，这种柔性表明随机突变引起剪接模式的改变是基因和生物进化的一种重要途径。

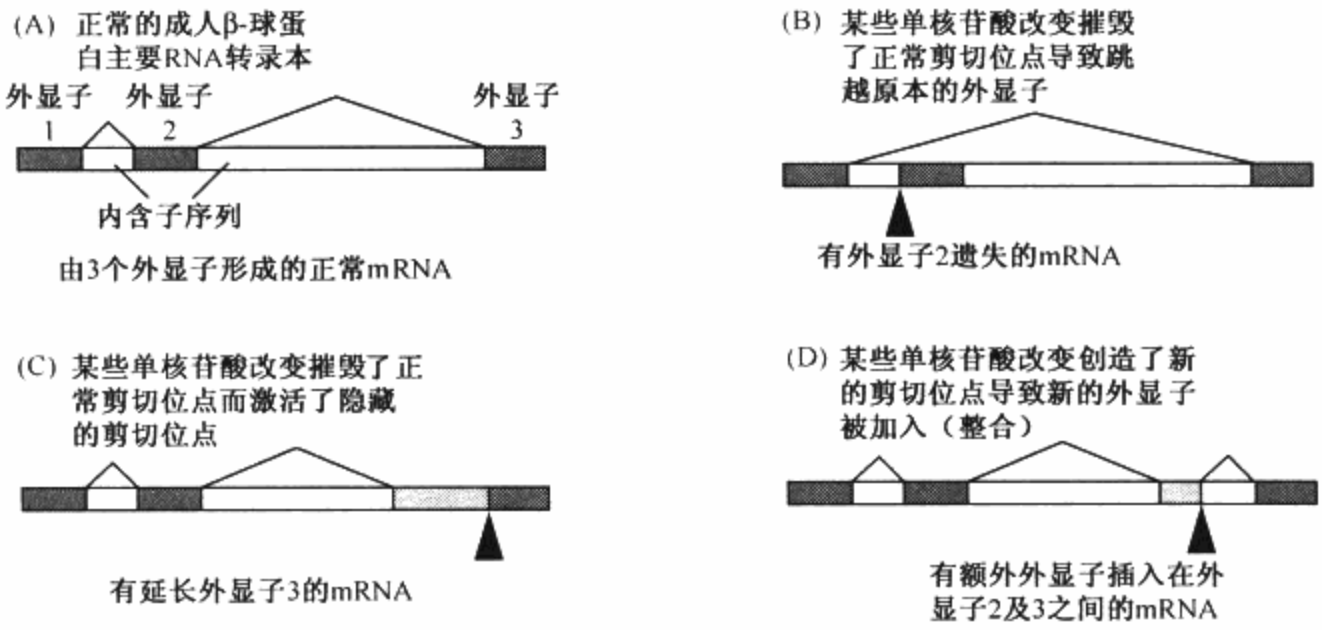


图 6-35　患有 β-地中海贫血病的人体中 β-珠蛋白初始 RNA 转录物的异常加工

在图中所示的例子中，该病由剪接位点的突变而引起，用黑色箭头表示。深蓝色框代表三个正常的外显子序列；红色直线用于表示在剪接 RNA 转录物时用到的核苷酸序列。淡蓝色框表示突变产生的最终 mRNA 中包含的新核苷酸序列。注意，当突变使正常剪接位点没有伴侣位点时，就会略过一个外显子，或者将附近一个或多个异常"隐蔽"剪接位点用作伴侣位点，如(C)和(D)中所示(摘自 S. H. Orkin, in The Molecular Basis of Blood Diseases[(G. Stamatoyannopoulos et al., eds.)], pp. 106～126, Philadelphia: Saunders, 1987)

RNA 剪接的柔性也意味着细胞能容易地调节 RNA 剪接模式。我们在本部分前面看到选择性剪接能从同一基因产生不同的蛋白质。有些选择性剪接的例子是组成型的；也就是说，一种生物的细胞连续合成经选择性剪接的 mRNA。然而在大多数情况下，剪接模式受到细胞调节，这样就能在不同的时刻和不同的组织内产生不同的蛋白质（图 6-27）。我们将在第 7 章再回到这个问题上来，讨论几个受调控的 RNA 剪接的特殊例子。

由剪接体催化进行的 RNA 剪接可能从自我剪接机制演变而来

最初发现剪接体的时候，它让分子生物学家产生迷惑。为什么 RNA 分子会代替蛋白质，在剪接位点识别和剪接化学反应中发挥重要的作用？为什么会利用一个套索中间产物，而不是利用看起来更简单的方式，即直接对它们进行切割和再连接后，只经过一步就将 5′和 3′剪接位点连接起来？这些问题的答案反映了人们所相信的剪接体的演变方式。

第 1 章简短地讨论过（并在本章最后一部分再次详细地讨论），有人认为早期细胞利用 RNA 而不是蛋白质作为它们主要的催化剂，并且它们将遗传信息储存在 RNA 而不是 DNA 序列中。RNA 催化的剪接反应可能在这些早期细胞中发挥重要作用。作为证据，许多自我剪接的 RNA 内含子（即没有蛋白质或任何其他 RNA 分子仍然能进行剪接的 RNA 中的内含子序列）目前依然存在。例如，在四膜虫的核内 rRNA 基因中、在少数 T4 噬菌体基因中，以及在某些线粒体和叶绿体基因中。

通过在试管中培育含有内含子序列的纯 RNA 分子，并观察剪接反应，可以鉴定出能自我剪接的内含子序列。用这种方法可以区分两类主要的自我剪接内含子序列。类型Ⅰ内含子序列通过将一个 G 核苷酸结合到内含子序列上，起始剪接反应；因而这个 G 被活化，形成一个攻击基团，它将打断剪接时切除的第一个磷酸二酯键（5′剪接位点的键）。在类型Ⅱ内含子序列中，一个活性特别高的 A 碱基是攻击基团，并形成了一个套索中间产物。其他情况下这

两类自我剪接内含子序列的反应途径都一样。可能它们两者都代表着非常古老的机制的残余（图 6-36）。

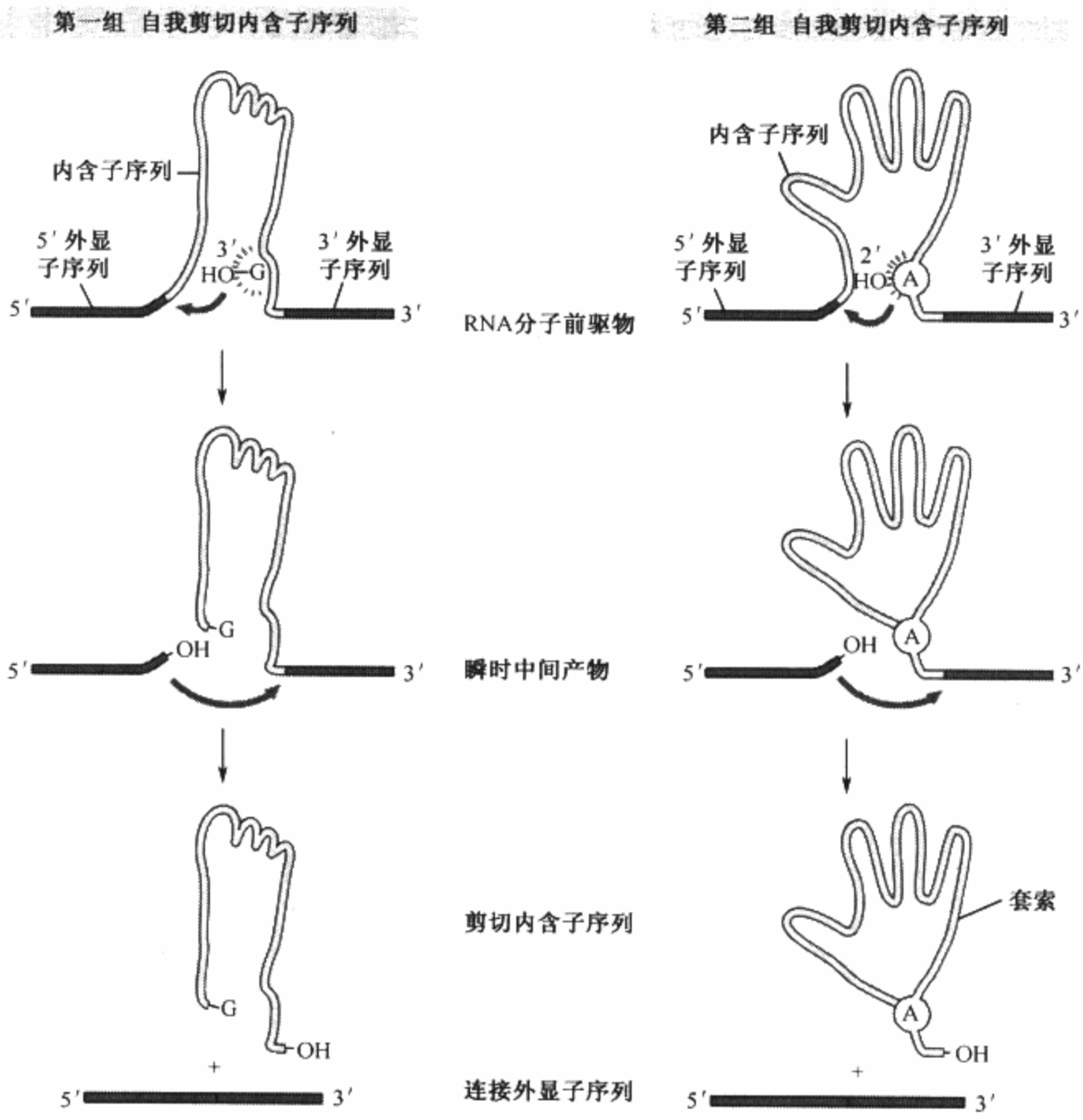


图 6-36　两类已知的自我剪接内含子序列

类型Ⅰ内含子序列结合一个游离的 G 核苷酸到 RNA 上的特殊位点来起始剪接，而类型Ⅱ内含子序列则使用内含子序列自身中的一个特别活跃的 A 核苷酸起始剪接。图中描绘这两个机制时强调了它们的相似性。两者通常都受到细胞内加速反应的蛋白质的帮助，但催化作用却是由内含子序列中的 RNA 介导的。这两种自我剪接反应需要将内含子折叠成高度特异的三维结构，提供该反应的催化活性（图 6-6）。类型Ⅱ内含子序列使用的机制以套索结构释放内含子，与剪接体催化的前体 mRNA 剪接途径非常相似（比较图 6-29）。真核细胞内大量 RNA 剪接由剪接体完成，自我剪接 RNA 代表少见的情况（摘自 T. R. Cech, *Cell* 44：207～210，1986）

内含子的核苷酸序列对于这两种类型的自我剪接反应是至关重要的；内含子 RNA 折叠成一个特殊的三维结构，这样就将 5′和 3′剪接衔接点并到一起，并提供精确定位的活性基团进行化学反应[图6-6(C)]。基于它们的剪接化学反应如此相似，有人提出剪接体的前体mRNA剪接机制是从类型Ⅱ剪接演化而来的。根据这个观点，当剪接体的 snRNP 取代类型Ⅱ内含子的结构和化学作用时，对内含子严格的序列限制可能就会消失，因而能够发生剪接的不同 RNA 的分子数量将大大增加。

RNA 加工酶产生真核 mRNA 3′端

前面解释过，RNA 聚合酶Ⅱ合成的前体 mRNA 的 5′端几乎在它刚从 RNA 聚合酶出来时发生帽化。然后，随着聚合酶沿基因继续移动，剪接体在 RNA 上组装起来，并标出内含子和外显子的边界。当 RNA 从聚合酶出现时，通过直接转移加帽和剪接组分到 RNA 上，RNA 聚合酶长长的 C 端尾协同了这些过程。正如我们在这一部分看到的，当 RNA 聚合酶在基因末端终止转录时，它使用类似的机制确保前体 mRNA 3′端得到正确加工。

可能有人会想到，最终是基因组中编码的信号指令形成了 mRNA 3′端（图 6-37）。这些 DNA 信号在 RNA 聚合酶移动经过它们时转录成 RNA，然后 RNA 由一系列 RNA 结合蛋白和 RNA 加工酶识别（作为 RNA）（图 6-38）。有两个多亚基蛋白质特别重要，它们分别叫做 CstF（切割刺激因子 F）和 CPSF（切割和多聚腺苷化特异因子）。这两种蛋白质都随 RNA 聚合酶尾巴一起移动，并在 RNA 分子刚从 RNA 聚合酶上出现时，转移到该 RNA 加工序列的 3′端。CPSF 的一些亚基与通用转录因子 TFⅡD 结合在一起，本章前面我们已看到，它参与了转录起始。在转录起始过程中，这些亚基可能从 TFⅡD 转移到 RNA 聚合酶尾巴上，并保持结合状态，直到聚合酶转录通过了基因末端。

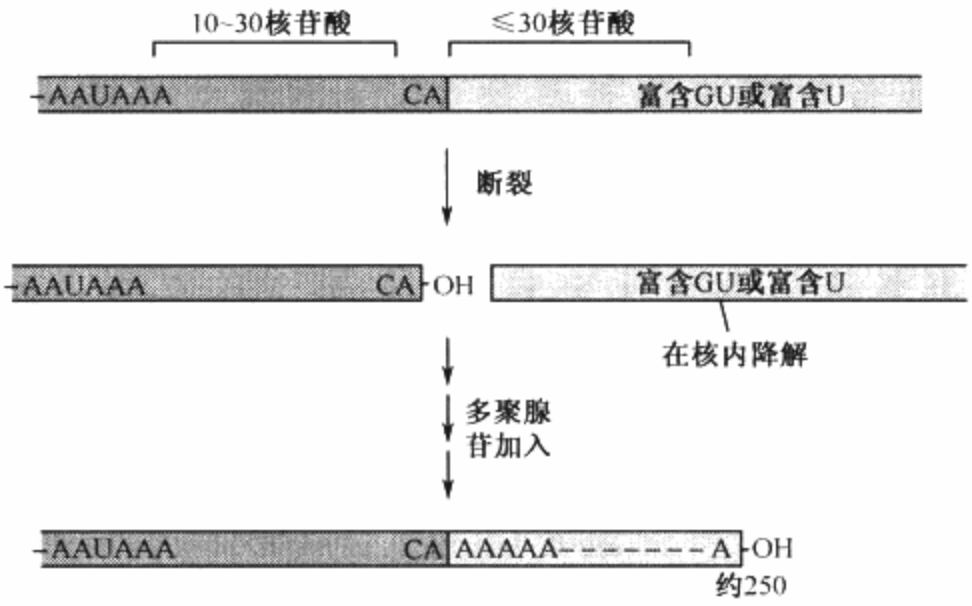


图 6-37　指导切割和多聚腺苷酸化以形成真核 mRNA 3′端的共有核苷酸序列

这些序列编码在基因组中并在转录成 RNA 后由特殊蛋白质识别。六聚体 AAUAAA 由 CPSF 接合，切割位点以外富含 GU 的元件由 CstF 结合（图 6-38），CA 序列则由切割步骤必需的第三种因子结合。与本章讨论的其他共有核苷酸序列（图 6-12）相似，图中所示的序列代表不同的单独切割和多聚腺苷酸化的信号

一旦 CstF 和 CPSF 结合到刚出现的 RNA 分子特定的核苷酸序列上，其他蛋白质就会和它们一起组装起来进行加工过程，生成 mRNA 3′端。首先对 RNA 进行切割（图 6-38）。然后一个称为多聚 A 聚合酶的蛋白质一次一个地把将近 200 个 A 核苷酸加到切割产生的 3′端上。这些加入核苷酸的前体是 ATP，并形成与普通 RNA 合成时一样的 5′→3′键（图 6-4）。与常见的 RNA 聚合酶不同，多聚 A 聚合酶不需要模板；因此真核生物 mRNA 的多聚 A 尾巴并不是直接编码在基因组中。一旦合成了多聚 A 尾巴，多聚 A 结合蛋白就在它上面组装，并以尚不了解的机制，决定了尾巴最终的长度。在 mRNA 从核内转移到细胞质期间，多聚 A 结合蛋白保持与多聚 A 尾巴的结合状态，并帮助在核糖体上合成蛋白质，这些我们在本章后面会看到。

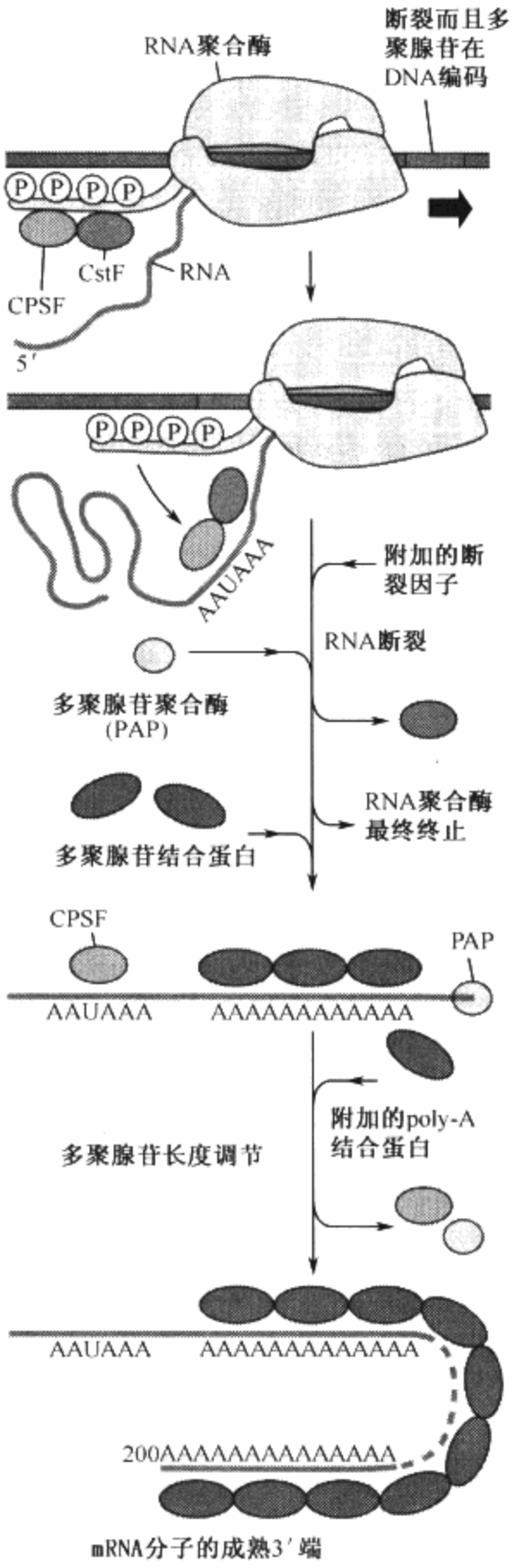


图 6-38　生成真核 mRNA 3′端的一些主要步骤

这个过程比细菌中的类似过程要复杂得多，在细菌中 RNA 聚合酶只是在终止信号处停下来，并释放其转录物 3′端和 DNA 模板（图 6-10）

真核生物前体 mRNA 3′端切除以后，RNA 聚合酶Ⅱ继续进行转录，在有些情况下，通过含有 3′切割位点信息的 DNA 以后，持续合成多达几百个核苷酸。但是不久聚合酶就松开它对模板的结合，并终止转录；切割位点下游的 RNA 片段然后就在细胞核内降解掉。还不清楚在 RNA 切割后是什么引起 RNA 聚合酶持续合成能力的丢失。一种观点认为，3′端加工因子从 RNA 聚合酶转移到 RNA 上引起 RNA 聚合酶发生构象变化，并松开与 DNA 的结合；另一种观点则认为刚从聚合酶合成出来的 RNA 5′端缺少帽子结构（和 CBC），在某种程度上为聚合酶终止转录提供了信号。

成熟的真核 mRNA 选择性地输出细胞核

现在我们已经看到，真核生物 mRNA 的合成和加工在细胞核内以有序的方式进行。然而，这些过程对真核细胞产生一个特殊的问题，尤其是那些所含内含子比外显子大得多的复杂生物。合成的前体 mRNA 中，只有一小部分（成熟 mRNA）对细胞有进一步的利用价值。其余的（切除的内含子、断裂 RNA 和异常剪接的前体 mRNA）不仅没有用处，而且如果不加以清除就会对细胞很危险。那么细胞是怎样区分它希望保留的较稀少的成熟 mRNA 分子和 RNA 加工产生的海量的废弃物呢？答案就在于从细胞核向细胞质（在这里合成蛋白质）转运 mRNA 的过程是高度选择性的，与正确的 mRNA 加工过程紧密偶联。这种偶联由核孔复合物完成，它只识别并转运完整的 mRNA。

我们已经看到，前体 mRNA 在合成和加工时，与许多蛋白质结合在一起，包括帽子结合复合体、SR 蛋白以及 poly（A）结合蛋白。在准备转运前，mRNA 分子似乎必须与合适的一组蛋白质结合，存在某些蛋白质，如帽子结合复合体，而缺少另一些，如 snRNP 蛋白。剪接期间置于 RNA 上的其他蛋白质似乎标记了外显子-外显子边界，因而代表剪接反应的完成。只有当合适的一组蛋白结合上 mRNA 之后，它才能引导通过核孔复合物，到达细胞质。正如第 12 章描述的，核孔复合物是核膜中的含水通道，直接连通核基质与细胞质。小分子（小于 5000Da）能自由扩散通过它们，然而细胞中大多数大分子，包括与蛋白质形成复合物的 mRNA，由于太大，如果没有特殊过程来转移它们，就不能通过核孔。在核孔内沿两个方向活跃地进行着物质转运。第 12 章

将会解释，大分子上的信号决定了它是从细胞核输出（如 mRNA 分子）还是输入（如 RNA 聚合酶）。对于 mRNA，标记剪接完成的结合蛋白特别重要，因为据我们所知它直接作为 RNA 输出因子（图 12-16）。很明显缺乏内含子的基因转录出的 mRNA，含有直接由其他 RNA 输出因子识别的核苷酸序列。因此真核细胞利用它们的核孔复合物作为大门，只让有用的 RNA 分子进入细胞质。

在前体 mRNA 分子刚从转录中的 RNA 聚合酶出现时就在它上面组装的所有蛋白质中，最丰富的是 hnRNP（核不均一核糖核蛋白）。这些蛋白质中有些（人类约有 30 种）从 RNA 上消除发卡螺旋，以便容易阅读 RNA 上的剪接和其他信号。另外一些则将通常在复杂生物基因中发现的特长内含子序列中的 RNA 包装起来（图 6-33）。除组蛋白之外，某些 hnRNP 蛋白是细胞核内含量最丰富的蛋白质，它们可能在区分成熟 mRNA 和加工废物过程中发挥特别重要的作用。hnRNP 颗粒（hnRNP 蛋白和 RNA 形成的类似核小体的复合物，（如图 6-33 所示）大都排斥在外显子序列以外，这可能是因为剪接体成分已经先结合上了。它们保留在切除的内含子上，可能作为它们停留在核内并最终降解的标记。

根据这个观点，内含子（与成熟 mRNA 一样）以结合蛋白的形式携带指令它们最终命运的信号。但是并不是全部 hnRNP 都停留在核内，据知有些 hnRNP 保持与完全加工的 mRNA 分子结合并伴随它们进入细胞质。可能是整套结合蛋白而不是单个蛋白质最终决定了一个 RNA 分子是否将离开细胞核。

对于昆虫巴尔比亚尼环基因而言，它的 mRNA 异常丰富，所以可以在电子显微镜下观察到 mRNA-蛋白质复合物从细胞核内输出。当这些基因转录后，就看到新形成的 RNA 被蛋白质（包括 hnRNP 和 SR 蛋白）包装起来。这种蛋白质-RNA 复合物发生一系列构型转变，可能反映了 RNA 加工过程，并直到形成弯曲的纤维为止（图 6-39）。然后这个弯曲纤维穿过核基质，进入核孔复合体（它的 5′端帽子在前），并且当它通过 NPC 时会再发生一系列其他的结构转变。这些以及其他一些观察到的现象表明，前体 mRNA-蛋白复合物和 mRNA-蛋白复合物是动态的结构，在 RNA 合成、加工、转运和翻译过程中它们会增加或丢失许多特定的蛋白质（图 6-40）。

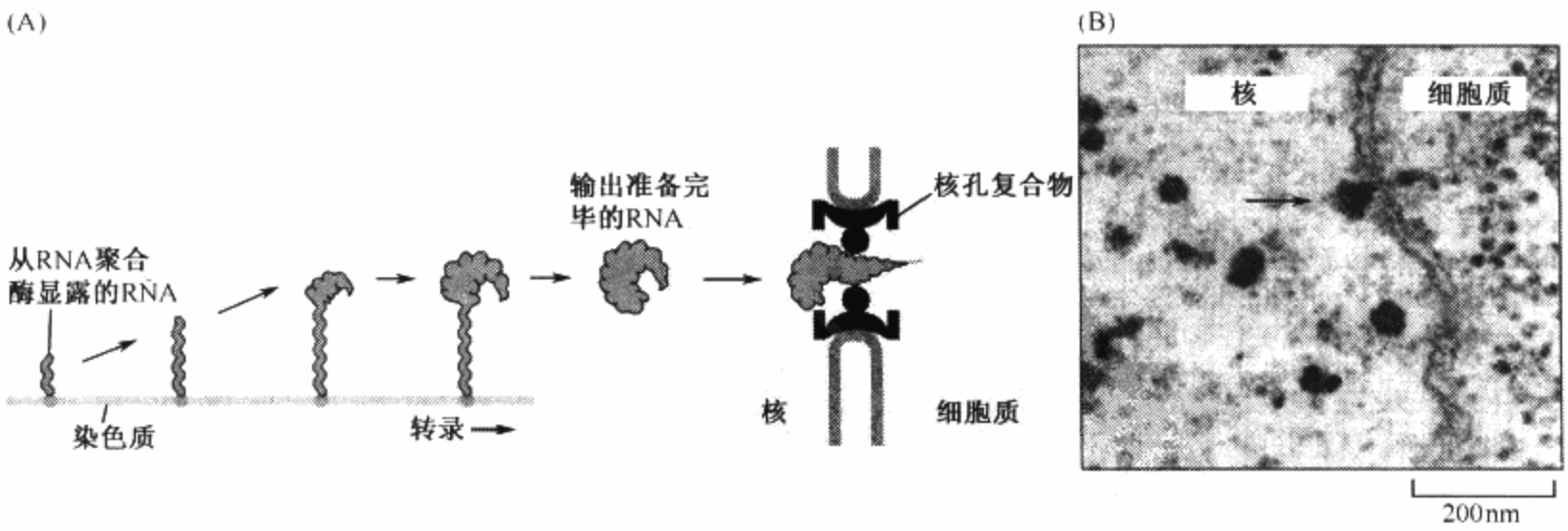


图 6-39　巨大的 mRNA 分子转运通过核孔复合物

（A）巴尔比亚尼环 mRNA 分子由 RNA 聚合酶合成，并包装成不同核蛋白的成熟过程。这幅昆虫细胞产生的异常丰余 RNA 的图是基于像（B）中这样的电镜显微照片。巴尔比亚尼环在第 4 章描述[(A)摘自 B. Danehlt, *Cell* 88：587～588，1997；(B) 摘自 B. J. Stevens and H. Swift, *J. Cell. Biol.* 31：55～77，1966 © The Rockefeller University Press]

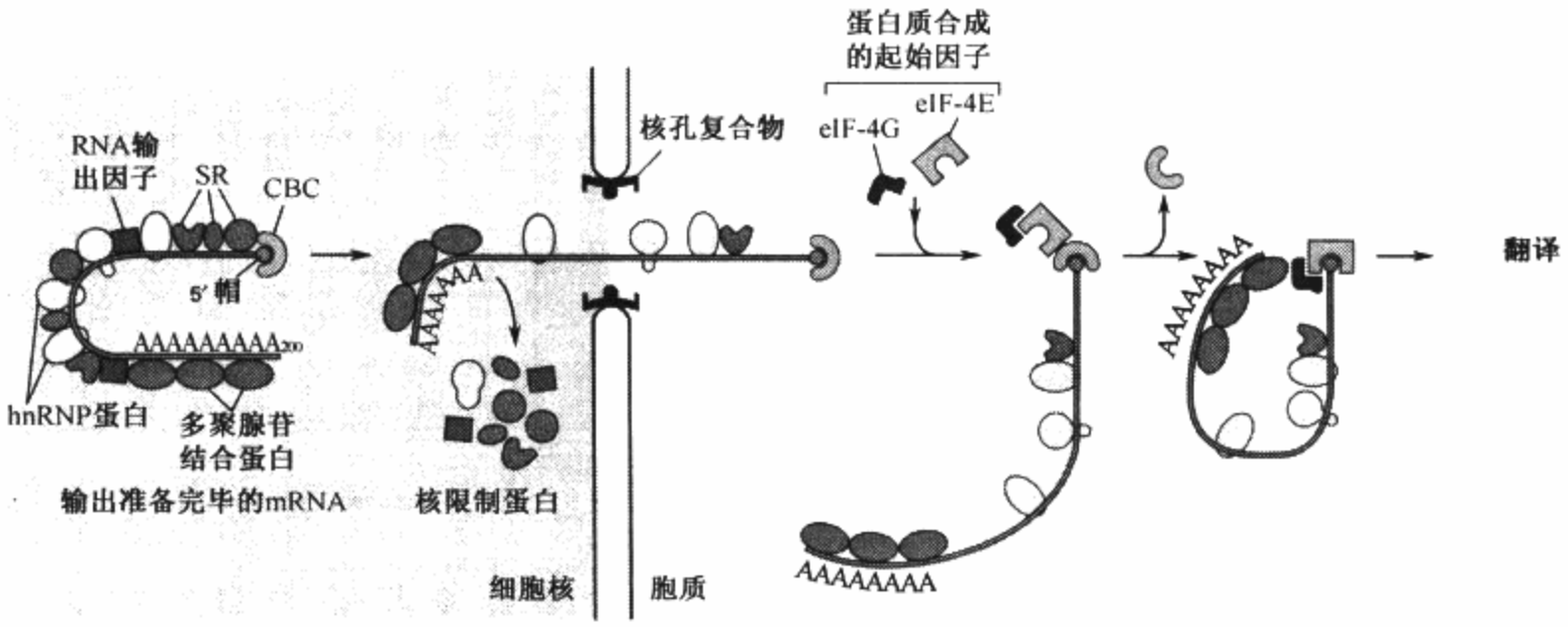


图 6-40　“准备输出”的 mRNA 分子及其转运通过核孔的简要说明

如图所示，当 mRNA 分子移动通过核孔时，有些蛋白质随同一起移动，而另一些保留在核内。一旦进入细胞质，mRNA继续脱去以前结合的蛋白质并获取新蛋白质；这些取代影响了随后的信息翻译。由于在核内与 mRNA 结合上的蛋白质中有些与 RNA 一起转运，所以它们能影响其随后在细胞质中的稳定性和翻译（图 12-16）。当剪接完成后有些这类蛋白质定位于外显子-内含子边界，因而代表了 RNA 中已正确剪接过的那些区域

在讨论 mRNA 离开核之后发生了什么之前，我们简短考虑一下非编码 RNA 分子的合成和加工是如何进行的。虽然也有许多其他的例子，但我们的讨论集中在对 mRNA 翻译成蛋白质至关重要的 rRNA 上。

许多非编码 RNA 也在细胞核内合成和加工

一个哺乳动物细胞内 RNA 占细胞干重的几个百分点：其中，只有 3%～5%是 mRNA。剩余的一部分代表尚未降解的内含子序列，但是细胞中大多数 RNA 行使着结构和催化功能（表 6-1）。细胞内含量最丰富的 RNA 是核糖体 RNA（rRNA），占迅速分裂细胞中 RNA 的近 80%。本章后面会讨论，这些 RNA 形成了核糖体的核心。与细菌（其中所有 RNA 都由一种 RNA 聚合酶合成）不同，真核生物有另外一个专门用于合成 rRNA 的酶，RNA 聚合酶Ⅰ。RNA 聚合酶Ⅰ在结构上与前面讨论的 RNA 聚合酶Ⅱ相似；然而，聚合酶Ⅰ缺少一个 C 端尾巴，这有助于解释它的转录产物为什么既不帽化也不多聚腺苷化。前面讨论过，这种差别帮助细胞区分非编码 RNA 和 mRNA。

由于每个 mRNA 分子经过多轮翻译，也能使蛋白质分子产量显著扩大，所以从那些单倍体基因组中只有一份拷贝的基因，可以合成出细胞内许多含量丰富的蛋白质。相比之下，核糖体的 RNA 成分是基因表达的终产物。一个生长中的哺乳动物，为了构建它的 1000 万个核糖体，在每个细胞世代各种核糖体 RNA 就必须合成将近 1000 万份拷贝。之所以能够合成足够数量的核糖体 RNA，只是因为细胞含有编码核糖体 RNA（**rRNA**）的 **rRNA 基因**的多份拷贝。即使是大肠杆菌也需要 7 份 rRNA 基因的拷贝，以满足细胞对核糖体的需要。人类每个单倍体细胞含有约 200 份 rRNA 基因拷贝，它们成小串地排列在 5 条不同染色体上（图 4-11），而非洲爪蟾的细胞每个单倍体基因组中含有约 600 份 rRNA 基因拷贝，它们位于一条染色体上的一个基因簇（图 6-41）。

图 6-41　电子显微镜中观察到串联排列的 rRNA 基因的转录

能容易地观察到交替转录的基因和非转录间隔子的模式。更高放大倍数的图像见图 6-9（摘自 V. E. Foe, *Cold Spring Harbor Symp. Quant. Biol.* 42：723～740，1978）

真核生物中存在 4 种 rRNA，每个核糖体都有它们的一份拷贝。其中三种 rRNA（18S、5.8S 和 28S）是通过化学修饰和切割单个巨大的前体 rRNA 后产生（图 6-42）；至于第 4 种（5SRNA）则由一种不同的聚合酶（RNA 聚合酶Ⅲ），从另外一串基因中合成出来的，且不需要经过化学修饰。为什么这个 RNA 单独合成还不清楚。

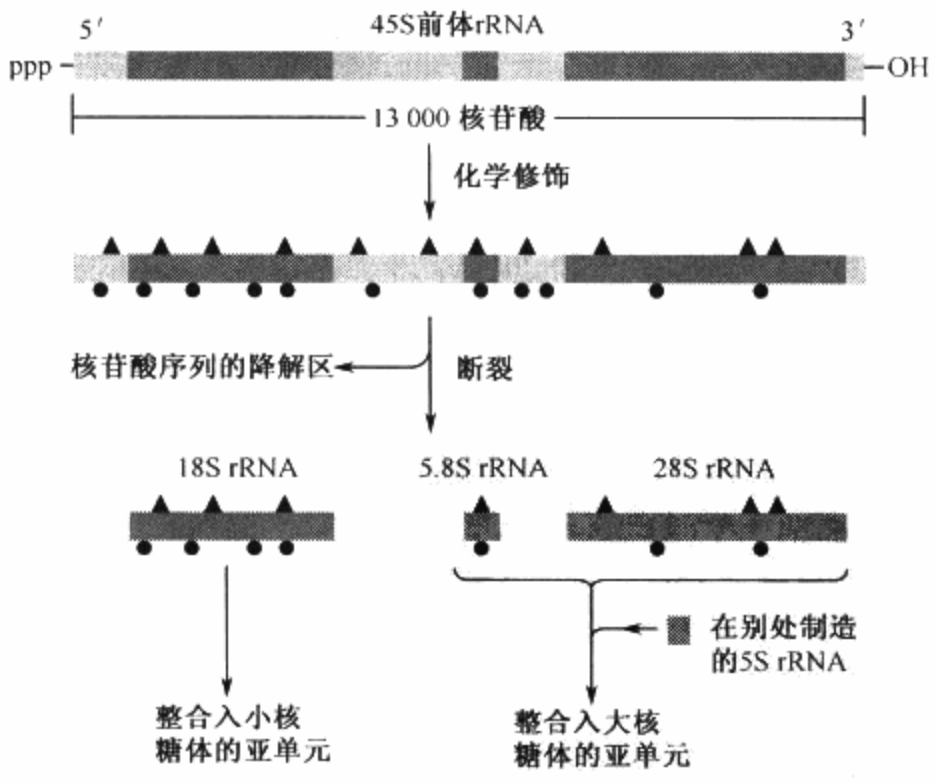
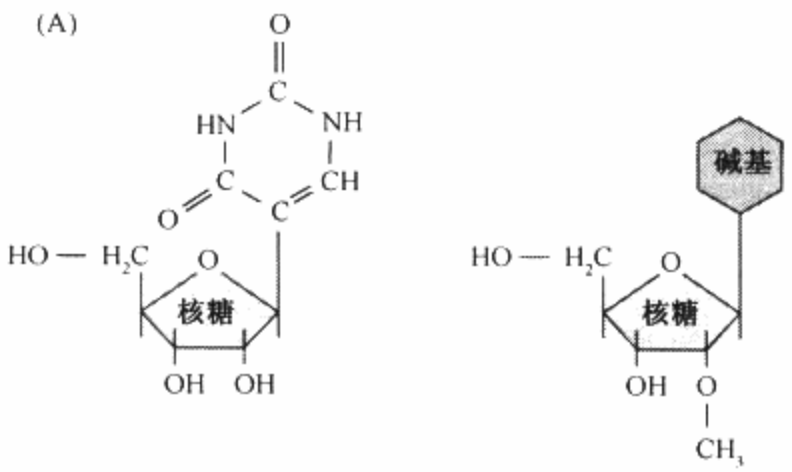


图 6-42　真核 45S 前体 rRNA 分子经化学修饰和核酸酶降解加工成三个单独核的糖体 RNA

如图所示，前体 rRNA 切割之前发生两种化学修饰（图 6-43）。这个前体 rRNA 中核苷酸序列几乎有一半被丢弃并在核内降解掉。rRNA 根据它们的“S”值命名，也就是指它超速离心时的沉降速率，S 值越大，rRNA 分子越大

前体 rRNA 长达 13 000 个核苷酸，从它切割出 rRNA 并组装成核糖体之前，它发生了广泛的化学修饰。这些修饰包括：核糖中约有 100 个 2′—OH 位置发生甲基化，还有约 100 个尿嘧啶核苷酸异构化为假尿嘧啶[图6-43(A)]。这些修饰的具体作用并不清楚，但是它们可能有助于最终 rRNA 折叠和组装，还可能引起核糖体功能发生微妙的变化。前体 rRNA 中每一种修饰都发生在特定位点上。这些位点由几百个“向导 RNA”规定，它们通过碱基配对将自身定位在前体 rRNA 上，因而将 RNA 修饰酶引导至合适的位点[图6-43(B)]。其他 RNA 可能通过在前体 rRNA 中引起构象变化，来促进前体 rRNA 切割成成熟的 rRNA。所有这些向导 RNA 都是一大类称作**核仁小 RNA**（或 **snoRNA**）中的成员，这样命名是因为这些 RNA 在细胞核内一个称作核仁的亚区室中发挥作用。许多 snoRNA 由其他基因的内含子编码，特别是那些编码核糖体蛋白质的基因。因此 snoRNA 由 RNA 聚合酶Ⅱ 合成，并从切割后的内含子序列中加工而来。



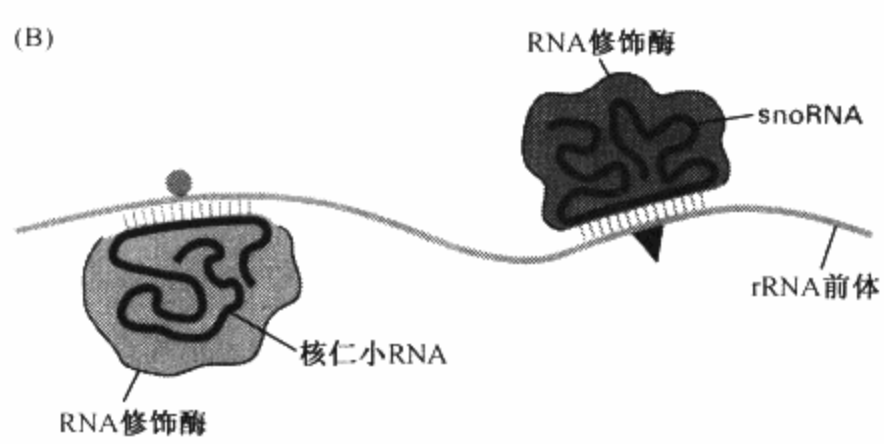


图 6-43　指导 RNA 对前体 rRNA 的修饰

（A）rRNA 合成之后发生的两种显著的共价修饰；红色原子表示与最初整合的核苷酸的差别。（B）如图所示，snoRNA 通过与前体 rRNA 上的互补序列碱基配对，定位于修饰位点。snoRNA 结合到蛋白质上，所得复合物称为 snoRNP。snoRNP 含有 RNA 修饰活性，可能是由于蛋白质的贡献，但也可能是由于 snoRNA 自身的贡献

核仁是生产核糖体的工厂

用光学显微镜观察真核生物细胞时，核仁是细胞核内看得最清楚的结构。因此早期的细胞学家对它的研究如此彻底，以至于 1898 年的一篇综述就能列出约 700 篇参考文献。现在我们知道了核仁是加工 rRNA 并将它们装配成核糖体的场所。与细胞内其他细胞器不同，它没

有膜的包被（图 6-44）；取而代之的是一个巨大的大分子聚集体，其中包括 rRNA 基因本身、前体 rRNA、成熟 rRNA、rRNA 加工酶、snoRNP、核糖体蛋白亚基和完成部分装配的核糖体。所有这些组分的密切联合，可以使核糖体的组装顺利进行。

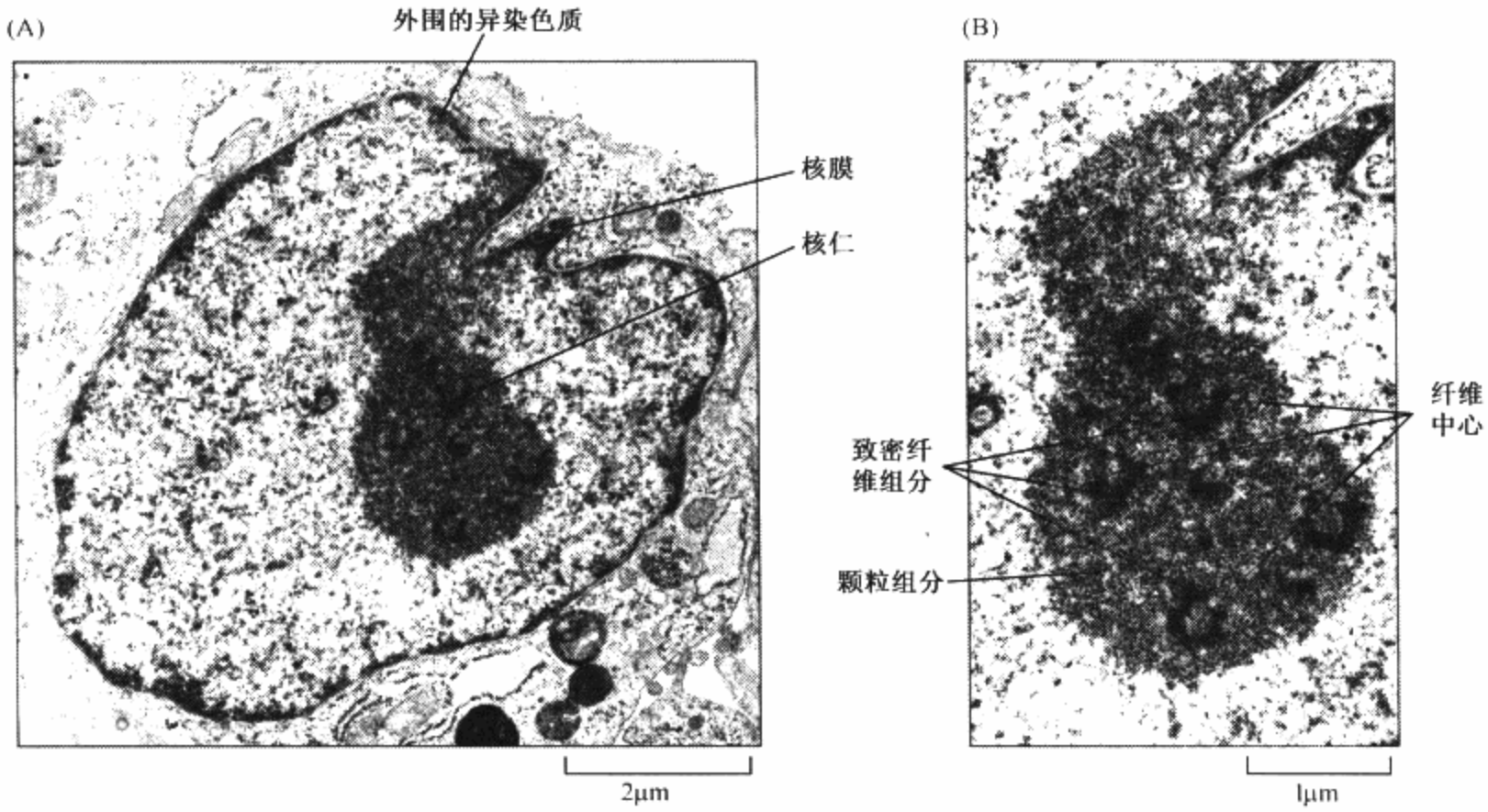


图 6-44　人类成纤维细胞核一小段截面的电镜显微照片，显示了其 3 个不同区域

（A）整个细胞核图片。（B）核仁的高倍放大图片。rRNA 基因转录据报道发生在核仁纤维中心和致密纤维组分之间，并且 rRNA 的加工以及组装成核糖体向外进行，即从致密纤维组分朝向周围的颗粒组分（经 E. G. Jordan 和 J. McGovern 允许后使用）

还不清楚细胞是怎样将核仁维系并组织起来的，但是不同的 RNA 分子在它的化学和结构中发挥中心作用，这表明核仁可能从细胞内一个古老而且主要由 RNA 催化的细胞器演化而来。在现今细胞中，rRNA 基因在核仁的形成过程也发挥重要作用。在人类二倍体细胞中，rRNA 基因分布在 10 个基因簇上，这些基因簇每个都位于具有两份拷贝的 5 条不同染色体的末端附近（图 4-11）。人类细胞每次发生有丝分裂时，染色体分离，核仁破裂；有丝分裂之后，当核仁重建时，10 条染色体又联合起来（图 6-45 和图 6-46）。由 RNA 聚合酶 I 进行的 rRNA 基因的转录对于这个过程是必需的。

可能有人会想到，核仁的大小反映了细胞正在生产的核糖体的数目多少。因此它的大小在不同细胞内变化很大，在单个细胞内也能发生改变，在那些蛋白质合成异常巨大的细胞内占据了核内容积的 25%。

图 6-47 显示了核糖体装配的简图。核仁除了在核糖体生物合成过程中发挥重要作用之外，还是其他 RNA 合成以及其他蛋白质-RNA 复合物组装的场所。例如，我们已经看到，在前体 mRNA 剪接中发挥作用的 U6 snRNP，由一个 RNA 分子和至少 7 种蛋白质组成（图 6-29）。U6 snRNA 在核仁内装配成 U6 snRNP 之前，受到 U6 snRNP 的化学修饰。其他重要的 RNA-蛋白质复合体，包括端粒（见第 5 章）和信号识别颗粒（我们将在第 12 章讨论），据报道也在核仁中组装。最后，为蛋白质合成转运氨基酸的 tRNA 也在核仁进行加工。因此，

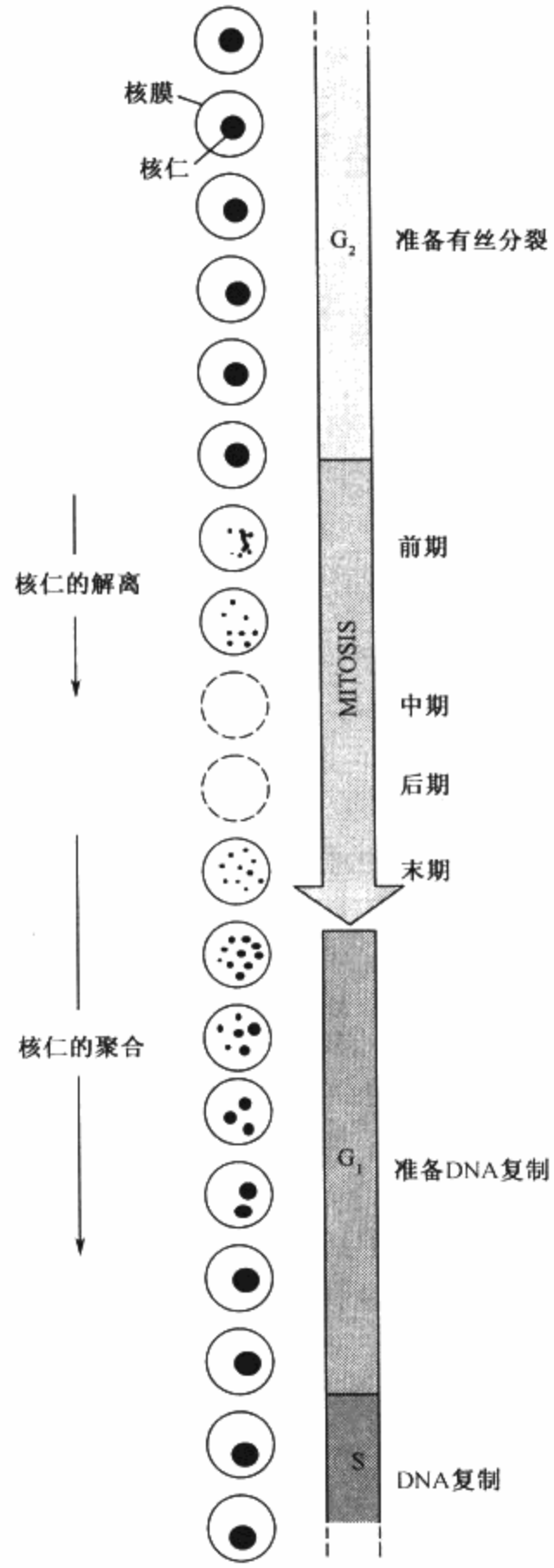


图 6-45　细胞周期中人类细胞的细胞核外观变化

图中只显示了细胞核。在大多数细胞中，有丝分裂期间，用间断的圆圈表示，核膜发生降解

可以认为核仁是一个巨大的工厂，在那里对许多非编码 RNA 进行加工，并将它们与蛋白质一起装配成多种核酸-蛋白复合物。

细胞核含有多种亚细胞核结构

虽然核仁是核内最显著的结构，但是也能观察到核内还存在多种其他核体结构，并对它们进行了研究（图 6-48）。这些结构包括 Cajal 小体（以 1906 年首次描述它们的科学家名字而

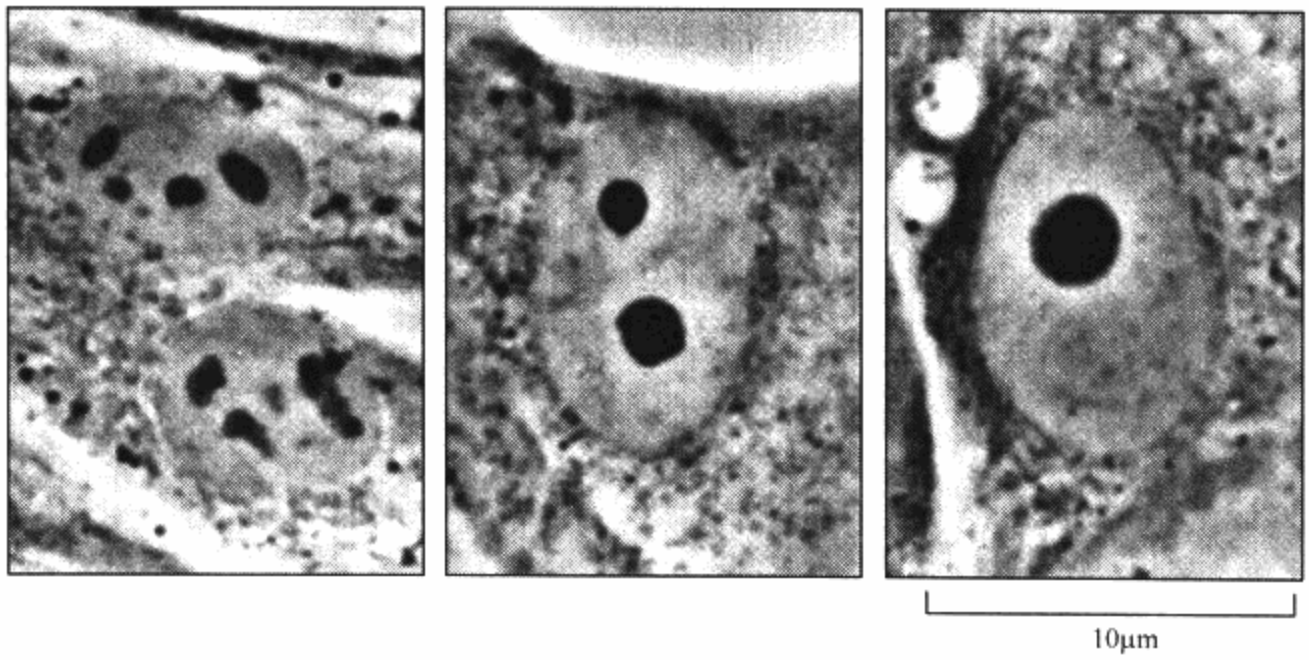


图 6-46　核仁融合

这些培养基中生长的人成纤维细胞光镜显微照片显示了核仁融合的不同阶段。有丝分裂以后，携带一簇 rRNA 基因的 10 条人类染色体各自都开始形成一个微小的核，但随着它们不断生长形成许多间期细胞典型具有的单个巨大细胞核时，它们就迅速合并（经 E. G. Jordan 和 J. McGovern 允许后使用）

命名）、GEMS（双卷曲小体），以及染色质间颗粒簇（也称作“斑点”）。与核仁类似，其他这些核内结构缺少膜且是高度动态的，它们的外形可能是蛋白质与 RNA（可能还有 DNA）组分紧密结合的结果，这些组分在那些参与基因表达的大分子合成、组装和储存中发挥作用。Cajal 小体和 GEMS 彼此相似，并在核内频繁配对；还不清楚它们是否真的是不同的结构。它们可能是 snRNA 和 snoRNA 进行最后的修饰并与蛋白质进行组装的场所。组成 snRNP 的 RNA 和蛋白质都在细胞质内进行部分组装，但是它们都转运到细胞核进行最后的修饰。有人提出 Cajal 小体/GEMS 也是剪接过程中重组完成以后循环利用 snRNP 并将其 RNA “复位”的场所。相比之下，有人认为染色质间颗粒簇是那些准备用于前体 mRNA 拼接过程的完全成熟的 snRNP 的储备（图 6-49）。

科学家一直难以搞清楚以上描述的亚核内小结构的功能。现在取得的大部分进展都依赖于遗传工具——检查设计的突变对小鼠的影响和自发突变对人类的影响。例如，GEMS 含有 SMN（运动神经元生存）蛋白。这种蛋白质编码基因的某些突变，引起了可遗传的脊髓肌肉萎缩，它是一种人类疾病，其特征就是肌肉逐渐消瘦。这种疾病似乎是因为 snRNP 组装以及随后进行的前体 mRNA 剪接存在轻微的缺陷而产生的。可想而知，更严重的缺陷将会是致命的。

考虑到亚细胞核区域在 RNA 加工中的重要性，可以推想到前体 mRNA 的剪接发生在核内某个特定的位置，因为剪接需要多种 RNA 和蛋白质组分。然而，我们已经看到，剪接组分在前体 mRNA 上的装配是共转录的；因此剪接必须在染色体上的许多位点进行。在第 4 章我们看到间期染色体占据了核内不连续的区域，而转录和前体 mRNA 剪接必须在这些区域内进行。然而，间期染色体它们本身是动态变化的，它们在核内确切位置与基因表达有关。例如，间期染色体的转录沉默区域常常与核被膜结合，据报道这里异染色质组分浓度特别高。当这些相同的区域开始活跃转录时，它们重新定位到核内侧位置，在这里富含合成 mRNA 所需要的成分。有人提出，一个典型的哺乳动物细胞尽管可能正按次序表达 15 000 个基因，转

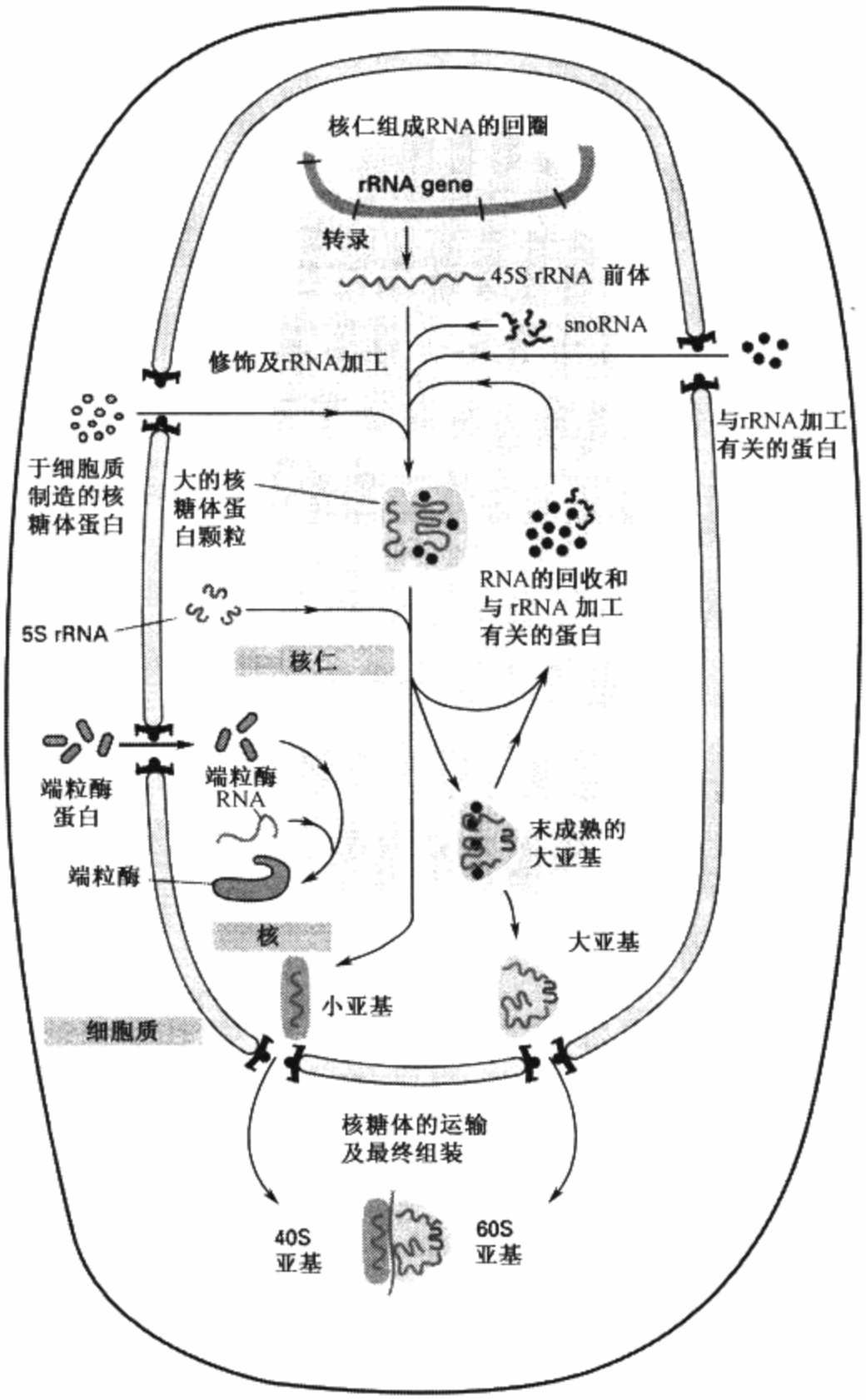


图 6-47　细胞核在核糖体和其他核糖核蛋白合成中的作用

45S 前体 rRNA 包装成一个巨大的核糖核蛋白颗粒，含有许多从细胞质输入的核糖体蛋白。尽管该颗粒保留在核内，但是随着它加工成未成熟的大小核糖体亚基时，会选择性地增加一些成分或去掉另一些成分。据认为两个核糖体亚基只有当它们各自独立地转运通过核孔进入细胞质时才获得最终的功能形式。其他核糖核蛋白复合体，包括这里所示的端粒，也在细胞核内组装

录和 RNA 剪接却可能只在核内几千个位点上进行。这些位点自身是高度动态变化的，并可能会引起转录和剪接组分联合起来形成一个小“装配线”，在这里这些组分的浓度会非常高。结果，细胞核似乎高度组织成多个亚区域，snRNP、snoRNP 和其他的核内组分根据细胞需要，以有序的方式在这些亚区域之间移动。

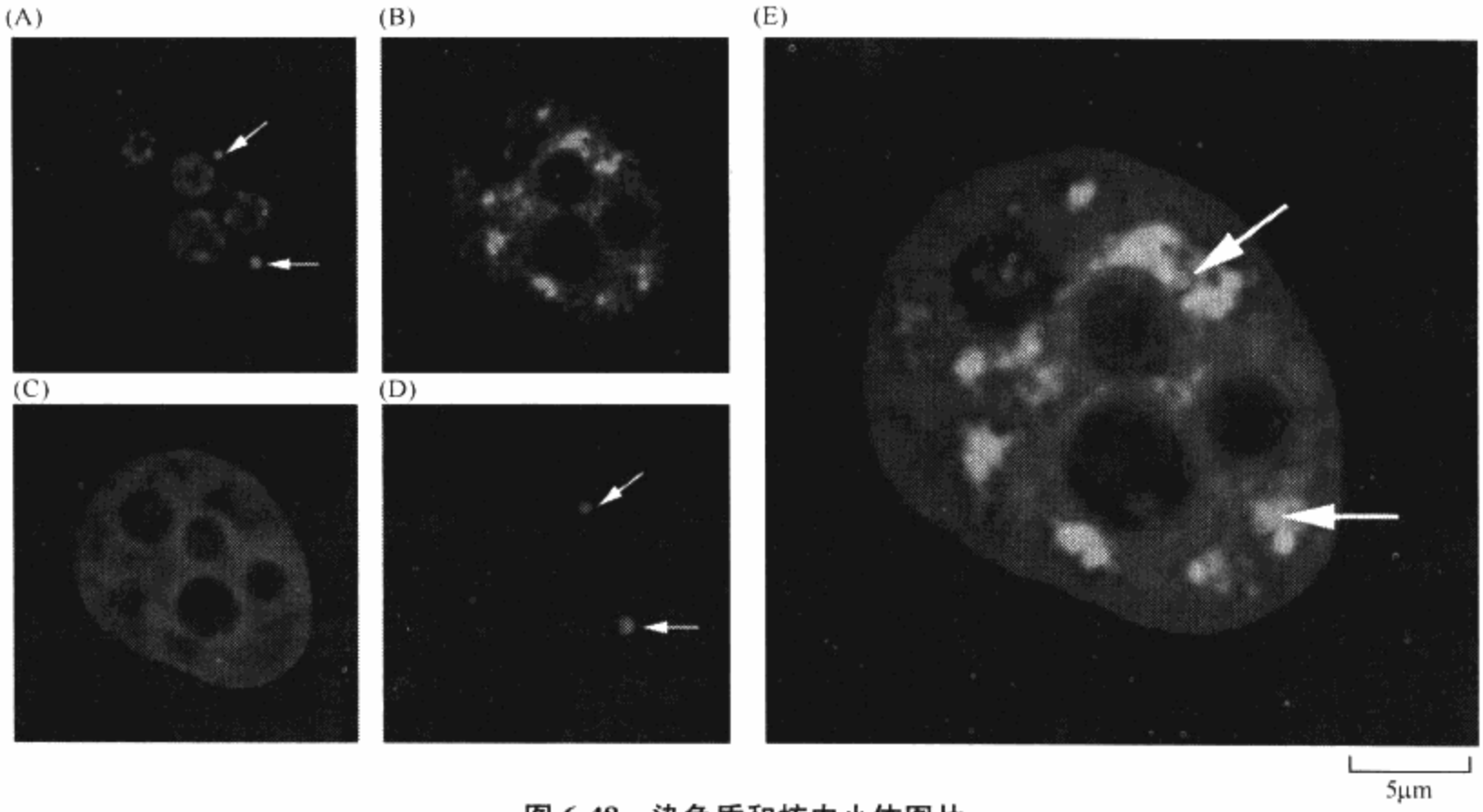


图 6-48　染色质和核内小体图片

（A）～（D）显示了相同人类细胞核的显微照片，这些细胞核每个都经过不同的处理，以显示特定的一组核结构。（E）显示全部 4 张图像叠加后放大的图像。（A）显示纤维蛋白的位置（多种 snoRNP 的成分），这种蛋白质存在于核仁和 Cajal 小体中，箭头表示后者。（B）利用参与前体 mRNA 剪接的一种蛋白质的抗体发现的染色质间颗粒簇或“斑点”。（C）染色后显示出巨大的染色质。（D）显示环蛋白的位置，这种蛋白质存在于 Cajal 小体中（箭头所指）（摘自 J. R. Swedlow and A. L. Lamond. *Gen. Biol.* 2：1～7，2001；显微照片经 Judith Sleeman 允许后使用）

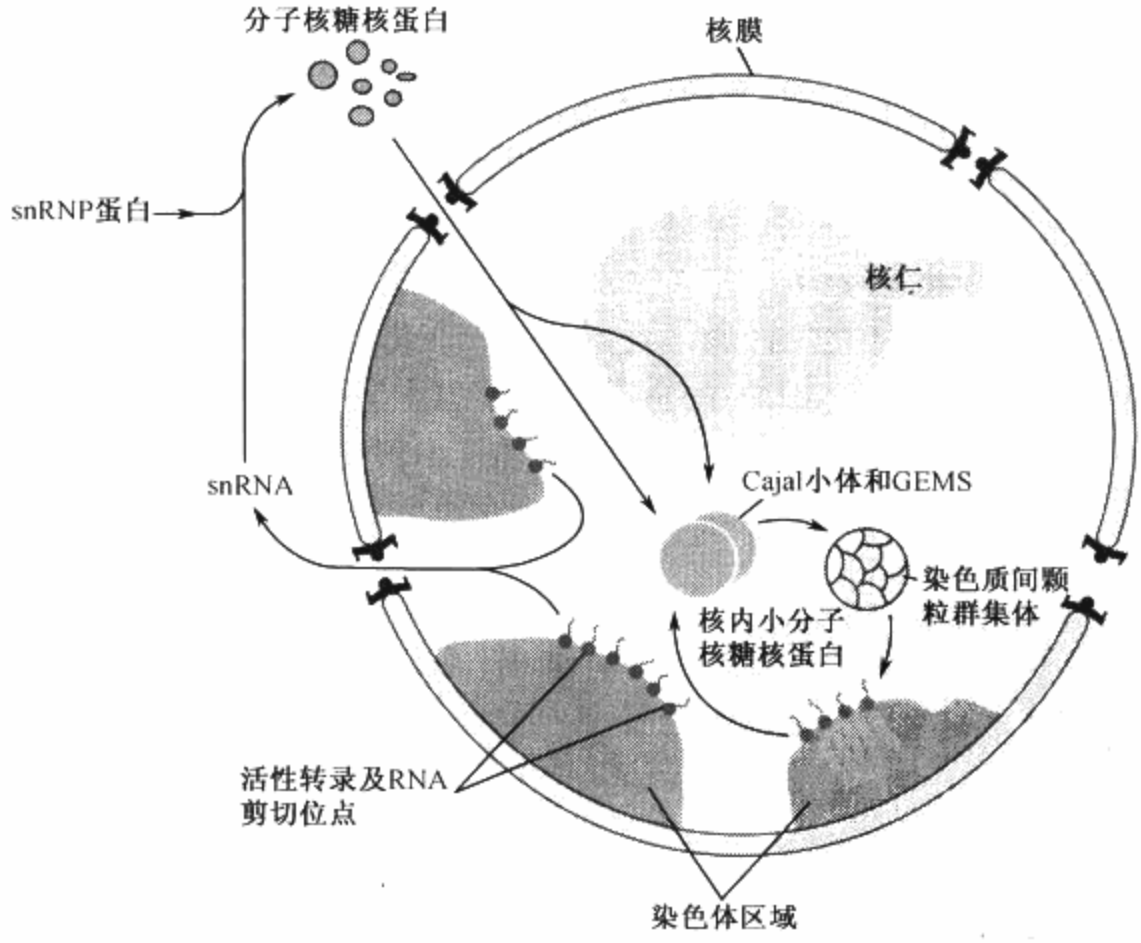


图 6-49　亚细胞核结构简图

典型的脊椎动物细胞含有 Cajal 小体，据报道它是 snRNP 和 snoRNP 发生最后修饰的地方。染色质间颗粒簇据报道是储存完全成熟的 snRNP 的位置。一个典型的脊椎动物细胞核拥有 20～50 个染色质间颗粒簇。snRNA 最初合成以后，从细胞核输出，在此之后它们发生 5′和 3′端加工并与 7 种常见的 snRNP 蛋白（称为 sm 蛋白）组装起来。这些复合物再重新转运到细胞核，snRNP 在 Cajal 小体中进行它们最终的修饰。此外，U6 snRNP 需要在核内由 snoRNA 进行化学修饰。活跃转录和剪接的位点（每个脊椎动物细胞核有将近 2000～3000 个位点）对应于电镜下观察到的“染色质周纤维”（摘自 J. D. Lewis and D. Tollervey，*Science* 288：1385～1389，2000）

小结

某一特定蛋白质开始合成之前，必须通过转录合成相应的 mRNA。细菌只有一种 RNA 聚合酶（将 DNA 转录成 RNA 的酶）。一分子 mRNA 的合成需要这种酶在一个启动子上起始转录，通过链的延伸合成 RNA，在终止子停止转录，并释放出 DNA 模板和互补的 mRNA 分子。在真核生物细胞内，转录的过程复杂得多，其中存在三种 RNA 聚合酶，分别命名为 RNA 聚合酶Ⅰ、Ⅱ、Ⅲ，在进化上它们彼此之间以及与细菌聚合酶之间都相关。

真核 mRNA 由 RNA 聚合酶Ⅱ合成。这种酶需要许多其他称为通用转录因子的蛋白质，来起始在纯化的 DNA 模板上的转录，还需要更多的蛋白质（包括染色质重建复合体和组蛋白乙酰基转移酶）来起始细胞内染色质模板上的转录。在转录延伸阶段，新生的 RNA 经过三种类型加工：5′端增加一个特殊的核苷酸（帽化），从 RNA 分子内部除去内含子序列（剪接），以及 RNA 3′端的生成（切割和多聚腺苷化）。这些修饰初期 RNA 转录物的 RNA 加工过程，有些（如涉及 RNA 剪接的过程）主要由特殊的小 RNA 分子进行。

对于某些基因，RNA 是最终的产物。在真核生物中，这些基因常常由 RNA 聚合酶Ⅰ或 RNA 聚合酶Ⅲ进行转录。RNA 聚合酶Ⅰ合成核糖体 RNA。这些核糖体 RNA 作为一个巨大的前体合成出来之后在核仁（一个与众不同的亚细胞核结构，也帮助加工细胞内一些较小的 RNA-蛋白质复合物）内进行化学修饰、切割并装配成核糖体。其他的亚细胞结构（包括 Cajal 小体和染色质间颗粒簇）是参与 RNA 加工的成分进行装配、储存和回收的场所。

从 RNA 到蛋白质

在前面的部分我们已看到，有些基因的终产物本身是 RNA 分子。例如，snRNP 和核糖体中存在的 RNA，然而细胞内大多数基因的成熟 mRNA 分子作为合成蛋白质途径的中间物。本部分我们考查细胞是怎样将携带在 mRNA 分子中的信号转变成蛋白质分子的。20 世纪 50 年代晚期，翻译这个技术最先吸引了生物学家的注意，当时把它作为“编码难题”提出来：RNA 线性核苷酸序列上的信息是怎样翻译到蛋白质中在化学上完全不同的亚单位——氨基酸的线性序列中去的？那时这个吸引人的问题在科学家当中激起了强烈的反响。这里存在一个由自然创立的密码，在进化超过 10 亿年后，最终还能由进化的产物之一——人类将其破解。实际上，不仅密码被一步一步破译，而且细胞阅读这个密码的复杂装置——核糖体，也最终在 2000 年从原子细节上得到了揭示。

mRNA 序列编码在一套核苷酸三联体中

一旦经转录和加工合成出 mRNA 之后，其核苷酸序列中的信息就用于合成蛋白质。转录作为信息转换的一种方式容易理解：因为 DNA 和 RNA 在化学结构上相似，所以 DNA 可以直接作为模板，通过互补碱基配对合成 RNA。正如转录这个词所代表的意思那样，就好像将手写的信息转变为打印的文本。语言本身和信息的修饰并没有改变，所使用的符号也紧密相关。

相比之下，将 RNA 中的信息转变成蛋白质就代表信息**翻译**成另一种语言，所使用的符号完全不同。而且，由于 mRNA 中只有 4 种不同的核苷酸，蛋白质中也只有 20 种不同的氨基酸，这种翻译不可能通过 RNA 中单核苷酸和蛋白质中单个氨基酸之间直接的一对一的对应关系来进行解释。基因中的核苷酸序列，经过 mRNA 介导，按规则翻译成蛋白质中的氨基

酸序列，这个规则称为**遗传密码**，它在 20 世纪 60 年代早期被破译。

mRNA 分子中的核苷酸序列以 3 个一组连续进行阅读。RNA 是 4 种不同核苷酸组成的线性多聚体，所以三核苷酸就可能存在 4×4×4＝64 种组合：三联体 AAA、AUA 和 AUG 等。但是，蛋白质中发现的氨基酸通常只有 20 种。要么有些核苷酸三联体从未得到利用，要么就是密码冗余，有些氨基酸由多个三联体指令。事实上第二种可能性是正确的，如图 6-50 中完全解密的遗传密码所示。RNA 中每一组连续的三核苷酸称为一个**密码子**，每个密码子或者规定一个氨基酸或者指令翻译过程终止。

	AGA									UUA					AGC				
	AGG									UUG					AGU				
GCA	CGA						GGA			CUA				CCA	UCA	ACA		GUA	
GCC	CGC						GGC		AUA	CUC				CCC	UCC	ACC		GUC	UAA
GCG	CGG	GAC	AAC	UGC	GAA	CAA	GGG	CAC	AUC	CUG	AAA		UUC	CCG	UCG	ACG		UAC	GUG	UAC
GCU	CGU	GAU	AAU	UGU	GAG	CAG	GGU	CAU	AUU	CUU	AAG	AUG	UUU	CCU	UCU	ACU	UGG	UAU	GUU	UGA
Ala	Arg	Asp	Asn	Cys	Glu	Gln	Gly	His	Ile	Leu	Lys	Met	Phe	Pro	Ser	Thr	Trp	Tyr	Val	终止
A	R	D	N	C	E	Q	G	H	I	L	K	M	F	P	S	T	W	Y	V	

图 6-50　遗传密码

每个氨基酸的标准单字母缩写位于其三字母缩写下方（见专题 3-1 中的每个氨基酸全称及其结构）。根据传统，书写密码子时，常常从 5′端核苷酸向左进行。注意，大多数氨基酸都由不止一个密码子代表，并且在指令每个氨基酸的一套密码子中有一些规律性。同一氨基酸的密码子倾向于在第一和第二个位置含有相同的核苷酸，而在第 3 个位置上不同。有 3 个密码子不编码任何氨基酸，而是作为终止位点，代表蛋白质编码序列的结束。有一个密码子——AUG，既是起始密码子，代表蛋白质编码信息的开始，又是指令甲硫氨酸的密码子

这种遗传密码在所有现今生物中得到普遍应用。尽管也在密码子中发现有少量微小的差异，它们主要存在于线粒体 DNA 中。线粒体拥有它们自己的转录和蛋白质合成系统，独立于细胞其余的系统运行，可以理解它们的小基因组能够容纳密码的较小改变（讨论见第 14 章）。

从理论上讲，根据解码过程开始位置的不同，一段 RNA 序列可以以三种不同的**可读框**中的任意一种进行翻译（图 6-51）。但是 mRNA 这三种可能的可读框只有一种编码了所需要的蛋白质，后面我们可看到在蛋白质合成开始时，位于每个 RNA 信息起始处的特殊标点信号是怎样设定正确可读框的。

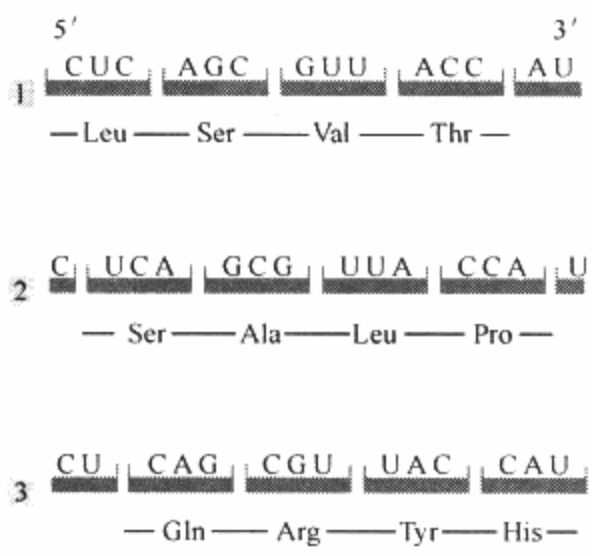


图 6-51　蛋白质合成中三种可能的可读框

在将一段核苷酸序列（蓝色）翻译成氨基酸序列（绿色）的过程中，mRNA 分子中的核苷酸序列以连续的三核苷酸从 5′端向 3′端阅读。因此从理论上讲，根据可读框的不同，同一的 RNA 序列能指令 3 个完全不同的氨基酸序列。然而，实际上这些可读框中只有一种含有真实的信息

tRNA 分子将氨基酸与 mRNA 中的密码子匹配

mRNA 分子中的密码并不直接识别它们指令的氨基酸：三联体核苷酸。例如，密码子不会直接与氨基酸结合，相反，mRNA 翻译成蛋白质时，有赖于衔接物分子，它们识别并结合到密码子上，而且在其表面的另一位点结合氨基酸。这些衔接物由一系列小 RNA 分子组成，称为**转移 RNA**（tRNA），每个长约 80 个核苷酸。

本章前面我们看到 RNA 分子能折叠成具有精确界定的三维结构。tRNA 分子就提供一个

显著的例子。折叠 tRNA 的 4 个短片段都是双螺旋，产生的分子如果画出轮廓，就像一片三叶草[图6-52(A)]。例如，一段多聚核苷酸链中的 5′-GCUC-3′序列可以与同一分子另一区域的 5′-GAGC-3′结合起来。三叶草形结构进一步折叠就形成一个紧密的 L 形结构，它通过分子不同区域之间形成的额外氢键维系在一起[图6-52(B),(C)]。

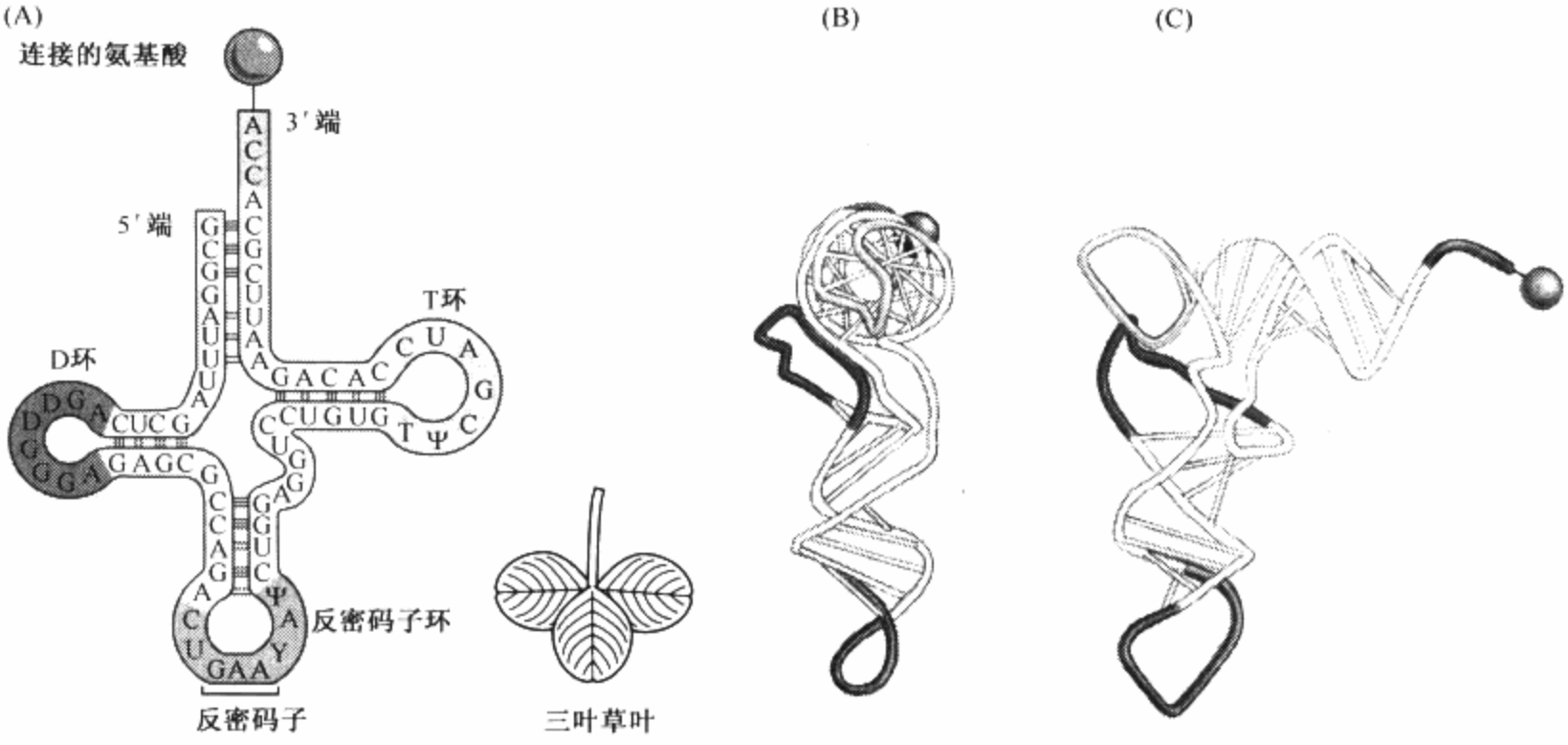


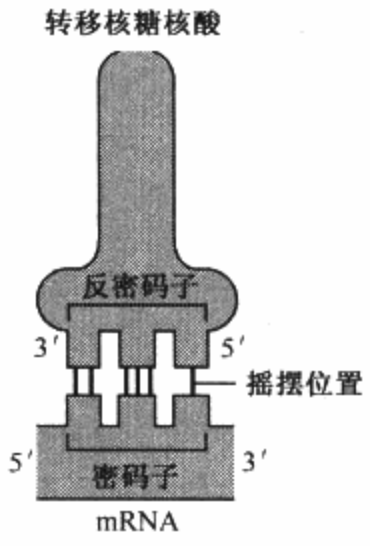
图 6-52　tRNA 分子

在这几幅图中，以不同的方式描绘了同一 tRNA 分子，本例中是苯丙氨酸（Phe）特异的 tRNA。（A）三叶草结构，常常用于显示产生分子内双螺旋区域的互补碱基配对（红线）。反密码子是与 mRNA 中的密码子碱基配对的三核苷酸序列。与密码子/反密码子配对匹配的氨基酸附着到 tRNA 的 3′端。tRNA 含有一些异常的碱基，它们在 tRNA 合成以后经过化学修饰而来。例如，ψ 表示的碱基（指假尿嘧啶，如图 6-43 所示）和 D 表示的碱基（二氢尿嘧啶，如图 6-55 所示）都来源于尿嘧啶。（B）和（C）基于 X 射线晶体衍射分析得到的实际存在的 L 形分子图形。虽然只描述了特定的一个 tRNA 分子——苯丙氨酸的 tRNA，但所有其他 tRNA 具有非常类似的结构。（D）分子的线性核苷酸序列。用彩色标记的核苷酸与（A）、（B）和（C）中的标记相对应

位于 L 形分子 2 个末端的区域中，未配对核苷酸对于 tRNA 在蛋白质合成中发挥作用很关键。其中一个区域形成**反密码子**，即与 mRNA 分子中互补密码子配对的一组三联体核苷酸；另一个是位于 tRNA 分子 3′端的一段短单链区域，这就是与密码子匹配的氨基酸结合到 tRNA 上的地方。

在前面的部分中，我们已经看到遗传密码是冗余的；也就是说几个不同的密码子可以指令同一个氨基酸（图 6-50）。这种冗余意味着对应每个氨基酸有不止一个 tRNA，或者有些 tRNA 分子能与多个密码子配对。实际上，这两种情况都存在。有些氨基酸拥有不止一个 tRNA，而有些 tRNA 结构要求在密码子前两个位置上碱基配对必须精确，第三个位置上则可以允许发生错配（或者摇摆）（图 6-53）。这种摇摆碱基配对解释了为什么同一氨基酸的这么多可选择密码子只在第三个核苷酸上存在差别（图 6-50）。在细菌中，摇摆碱基配对使得仅用 31 种 tRNA 分子就能将 20 种氨基酸与它们的 61 个密码子匹配起来。可是不同物种之间，

不同类型的 tRNA 确切数目不同。例如，人类有 497 个 tRNA 基因，但它们中间只代表了 48 种不同的反密码子。



细菌

摇摆密码子碱基	可能的反密码子碱基
U	A, G, or I
C	G, or I
A	U, or I
G	C, or U

真核生物

摇摆密码子碱基	可能的反密码子碱基
U	G, or I
C	G, or I
A	U
G	C

图 6-53　密码子与反密码子之间的摇摆碱基配对

如果第一排所列的核苷酸位于密码子第 3 位或摇摆位置，那么它就能与第二排所列核苷酸中任何一个配对。因而，比如说，当 tRNA 反密码子的摇摆位置存在次黄嘌呤（I）时，那么这个 tRNA 就能识别细菌的 3 个不同密码子中的任何一个，或者真核生物的 2 个密码子中的任何一个。tRNA 中的次黄嘌呤通过鸟嘌呤脱氨基形成（图 6-55），这种化学修饰在tRNA合成以后进行。非标准碱基配对，包括与次黄嘌呤形成的配对，通常较常规碱基配对要弱一些。注意密码子-反密码子碱基配对在密码子第 1 位和第 2 位更严谨一些：这里只允许存在常规的碱基对。细菌和真核生物的摇摆碱基配对相互作用之间的差别，可能是因为它们核糖体结构的细微差别，这些核糖体是合成蛋白质的分子装置（摘自 C. Guthrie and J. Abelson, in the Molecular Biology of the Yeast *Saccharomyces*: Metabolism and Gene Expression, pp. 47～528. Cold Spring Harbor Laboratory Press, 1982）

tRNA 从细胞核输出前受到共价修饰

我们已经看到，大多数真核生物 RNA 被允许从细胞核输出之前都发生了共价改变，tRNA 也不例外。真核 tRNA 由 RNA 聚合酶Ⅲ合成。无论细菌还是真核生物的 tRNA 通常都以巨大的前体 tRNA 合成出来，然后将它们截短，产生成熟的 tRNA。此外，有些 tRNA 前体（来自于细菌和真核生物）含有必须剪接掉的内含子。这个剪接反应在化学上与前体 mRNA 的剪接反应不同，tRNA 剪接通过一种由蛋白质催化的剪切-粘贴机制进行，而不会产生套索中间产物（图 6-54）。截短和剪接过程都要求前体 tRNA 正确地折叠成三叶草构型。因为错误折叠的 tRNA 前体不能进行正确的加工，所以认为截短和剪接反应是合成 tRNA 过程中的一个质量控制步骤。

所有的 tRNA 也发生各种化学修饰，每个成熟 tRNA 分子 10 个核苷酸中就约有一个是由标准的 G、U、C 或 A 核糖核苷酸改变而来。tRNA 存在超过 50 种不同的修饰，图 6-55 显示了其中一些。有些经过修饰的核苷酸（最著名的是次黄嘌呤核苷，由鸟嘌呤脱氨基后而来）

图 6-54　停留在前体 tRNA 上的 tRNA 剪接内切核酸酶的结构

内切核酸酶（一种四亚基酶）去除 tRNA 内含子（蓝色），然后另一种酶，一种多功能 tRNA 连接酶（未显示），将两段 tRNA 连接起来（经 Hong Li, Chistopher Trotta 和 John Abelson 允许后使用）

影响反密码子的构型和碱基配对，因而促进了 tRNA 分子识别合适的 mRNA 密码子（图 6-53）。其他一些修饰核苷酸则影响 tRNA 与正确氨基酸结合的精确性。

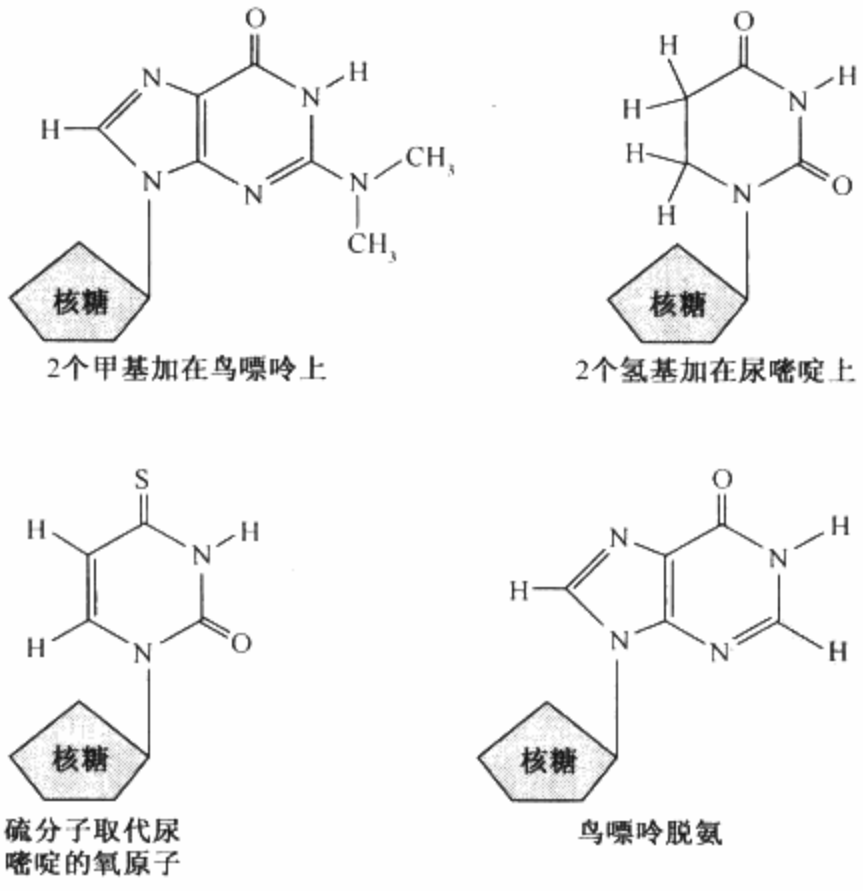


图 6-55　tRNA 分子中发现的一些异常核苷酸

这些核苷酸由正常核苷酸整合到 RNA 链以后经共价修饰产生。大多数 tRNA 分子中大约 10%的核苷酸都发生修饰（图 6-52）

特异的酶将每个氨基酸与它合适的 tRNA 分子偶联起来

我们已经看到，为了阅读 DNA 中的遗传密码，细胞生成了一系列不同的 tRNA。现在我们考查每个 tRNA 分子是怎样与 20 种氨基酸中合适的伴侣偶联起来的。识别和结合正确的氨基酸依赖于**氨酰-tRNA 合成酶**，它将每个氨基酸与其适当的一组 tRNA 分子连接起来（图 6-56 和图 6-57）。对于大多数细胞而言，相应于每一种氨基酸存在一个不同的合成酶（也就是说，总共有 20 种合成酶）；一种合成酶将甘氨酸结合到识别甘氨酸密码子的所有 tRNA 上，另一种则将丙氨酸结合到识别丙氨酸密码子的所有 tRNA 上，依此类推。然而许多细菌含有的合成酶不到 20 种，相同的合成酶负责将不只一种氨基酸连接到合适的 tRNA 上。在这些情况下，一个合成酶将相同氨基酸放置到两个不同的 tRNA 上，只有其中一个拥有与这个氨基酸匹配的反密码子，然后另一种酶对每个“错误”结合的氨基酸进行化学修饰，这样就能与共价相连的 tRNA 上的反密码子匹配起来。

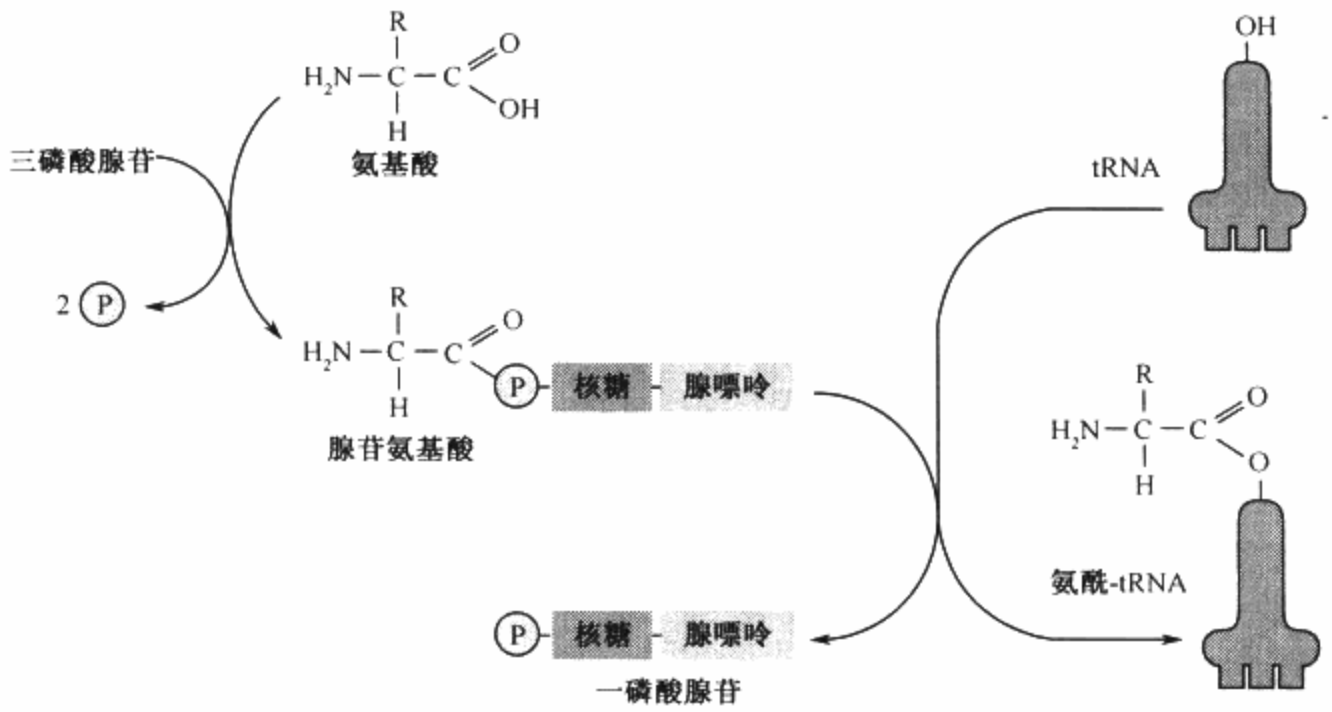


图 6-56　氨基酸活化

图中显示了两步反应过程，在此反应中氨基酸（其侧链用 R 表示）由氨酰-tRNA 合成酶活化后用于蛋白质合成。如图所示，ATP 水解释放的能量用于将每个氨基酸以高能键连接到它的 tRNA 分子上。首先，氨基酸通过其羧基直接与另一半的 AMP 相连，形成腺苷氨基酸而活化；与 AMP 相连通常是一个能量上不利的反应，因而由贡献 AMP 的 ATP 分子水解推动。AMP 连接的氨基酸的羧基然后转移到 tRNA 分子 3′端核糖的羟基上，而不必离开合成酶。这个转移过程用活化的酯键将氨基酸和 tRNA 连接起来，并形成最后的氨酰-tRNA 分子。图中未显示合成酶

合成酶催化将氨基酸连接到 tRNA 3′端的反应是细胞内许多偶联 ATP 水解放能的反应之一，该反应在 tRNA 和氨基酸之间形成一个高能键。这个键的能量用于蛋白质合成后期将氨基酸共价连接到生长的多肽链上。

虽然 tRNA 分子在将核苷酸序列转变成氨基酸序列过程中充当最后的转换器，但是在解码过程中，氨酰-tRNA 合成酶也是同样重要的转换器（图 6-58）。这个事实通过一个精巧设计的实验得到证明，该实验中，一个氨基酸（半胱氨酸）已经结合到特异的 tRNA 上以后，通过化学方法转变成另一个氨基酸（丙氨酸），当这种“杂交”氨酰-tRNA 分子用于无细胞系统中的蛋白质合成时，在蛋白质链中每一个会利用该 tRNA 的地方都会插入这个错误的氨基酸。虽然细胞拥有几种质量控制机制来避免这种错误，但这个实验清楚地表明，遗传密码

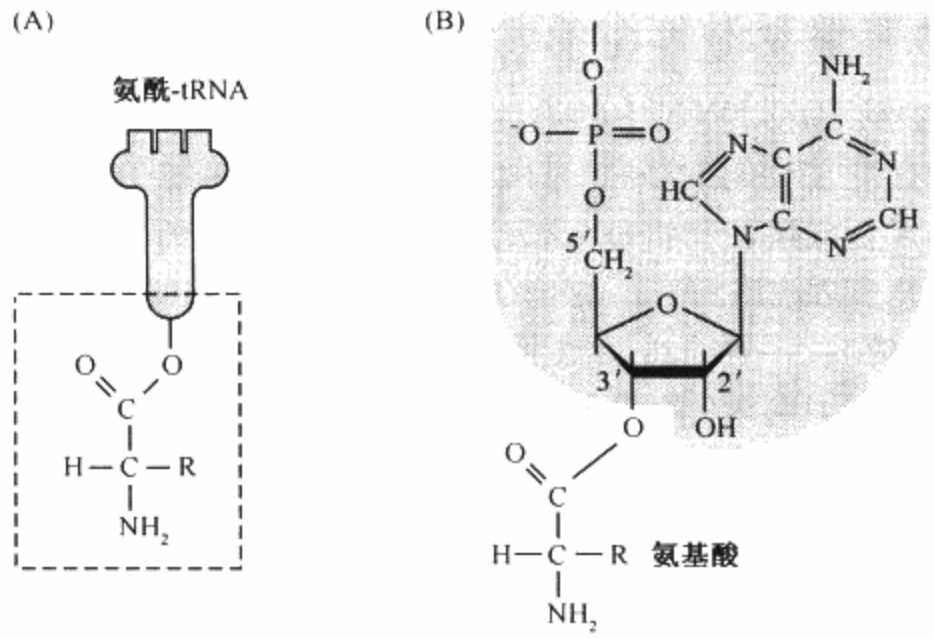


图 6-57　氨酰-tRNA 连接的结构

氨基酸的羧基端与核糖形成酯键。由于该酯键的水解伴随大量有利的自由能变化，所以通过这种方式维持的氨基酸称为被活化。（A）结构简图。氨基酸连接到 tRNA 3′端核苷酸上（图 6-52）。（B）对应（A）中方框区的实际结构。存在两类主要的合成酶：一类将氨基酸直接连接到核糖的 3′—OH 上，另一类首先将它连接到 2′—OH 上。在后一种情况下，随后发生转移反应将氨基酸转移到 3′位置。如图 6-56 所示，“R 基团”代表氨基酸的侧链

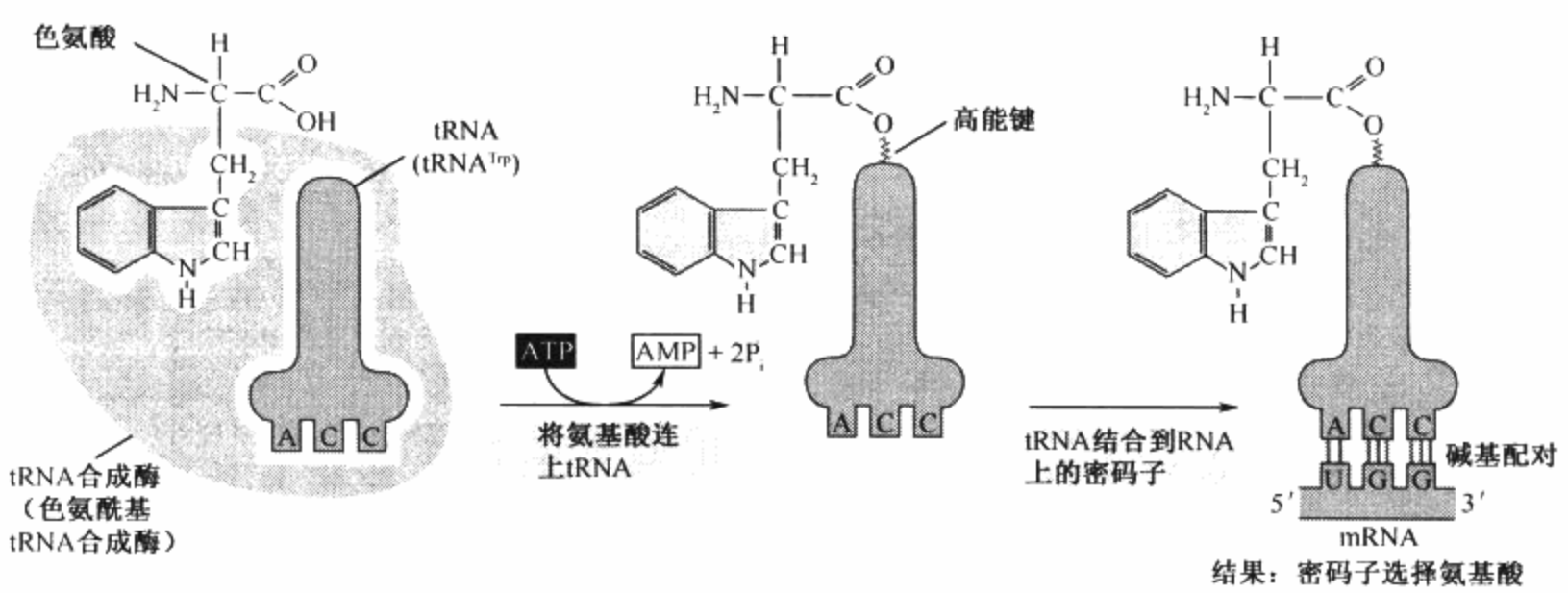


图 6-58　通过两个适配器先后作用的方式翻译遗传密码

第一个适配器是氨酰-tRNA 合成酶，它将特定的氨基酸与其对应的 tRNA 连接起来；第二个适配器是 tRNA 分子自身，其反密码子与 mRNA 上合适的密码子形成碱基配对。这两个步骤任意一步出现错误都导致将错误的氨基酸整合到蛋白质链中。在图中所示的一系列事件中，由 mRNA 上 UGG 密码子选择出色氨酸

由两套连续发挥作用的转换器进行翻译。这两个转换器中每个都使一个分子表面与另一个分子高度特异地结合，也正是由于它们的联合作用才将 mRNA 分子中每三个核苷酸序列，即每个密码子，与其专门的氨基酸结合起来。

RNA 合成酶进行的编辑确保了精确性

许多机制共同作用确保了 tRNA 合成酶将正确的氨基酸同每个 tRNA 连接起来。合成酶必须首先筛选出正确的氨基酸，大多数都靠一种包括两个步骤的机制完成。首先，正确的氨

基酸由于具有与其合成酶的活性位点口袋最高的亲和力，因此会优先于其他 19 种氨基酸结合到活性位点。特别是，会有效地排斥出比正确氨基酸更大的氨基酸。然而，如果只是通过一步识别的机制，就很难精确地区分出两个近似的氨基酸，如亮氨酸和缬氨酸（它们只有一个甲基的差别）。在氨基酸共价连接到 AMP 之后便进行第二步识别（图 6-56）。tRNA 结合上合成酶时，迫使氨基酸进入合成酶上的另一个口袋，它的精确大小正好将正确氨基酸排斥出去，却允许密切相关的氨基酸进入。一旦有一个氨基酸进入这个编辑口袋，它就从 AMP 上水解（或者如果氨酰-tRNA 键已经形成的话就从 tRNA 自身上水解）并从酶上释放下来。这个水解编辑与 DNA 聚合酶进行的编辑类似（图 6-59），它使 tRNA（负载）的整体精确性提高到将近每 40 000 次偶联只发生一次错误。

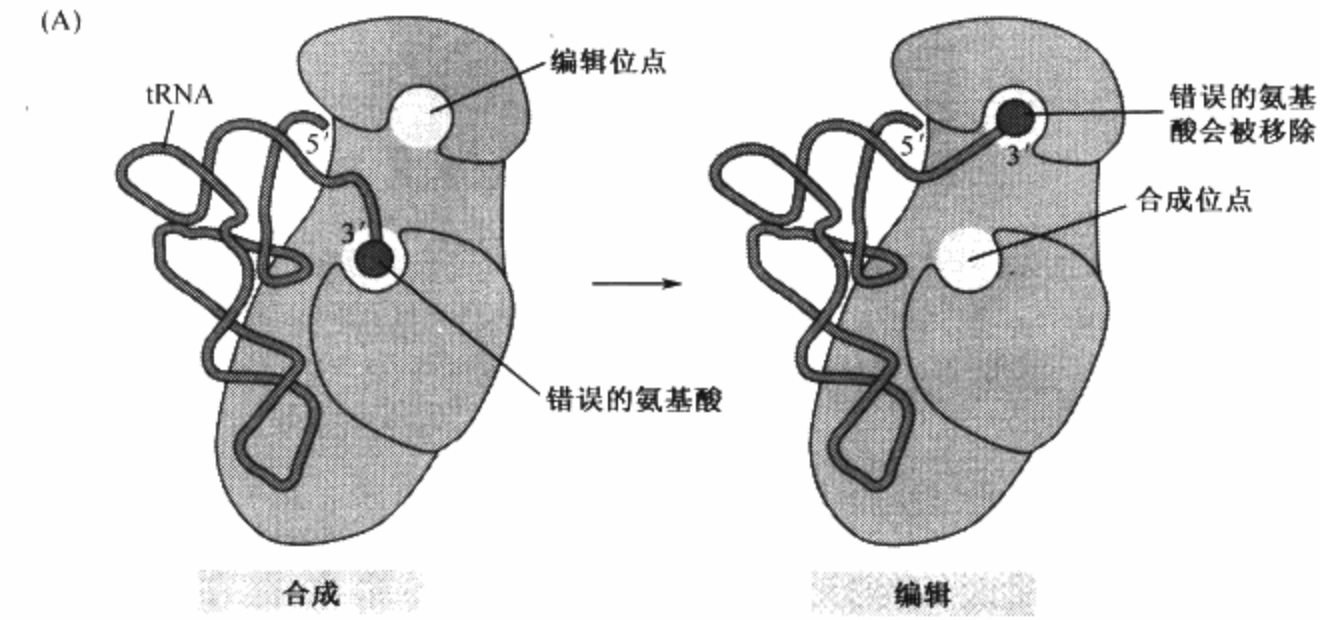


图 6-59　水解编辑

（A）tRNA 合成酶通过水解编辑错误附着的氨基酸来去除它们自身的偶联错误。如文中所述，编辑位点拒绝正确的氨基酸。（B）DNA 聚合酶的校错过程显示出一些相似性；然而，由于去除过程强烈依赖于与模板的错配，所以它又有很大不同

tRNA 合成酶也必须识别正确的一组 tRNA，合成酶与 tRNA 之间广泛的结构和化学互补性使酶能感受到不同特征的 tRNA（图 6-60）。大部分 tRNA 合成酶直接识别匹配的 tRNA 反密码子。这些合成酶含有 3 个相邻的核苷酸结合口袋，其中每个的形状和电荷与反密码子的互补。对于其他的合成酶，受体的主干核苷酸序列才是决定识别的关键因素。然而，在大多数情况下，tRNA 上许多位置的核苷酸都由合成酶进行“阅读”。

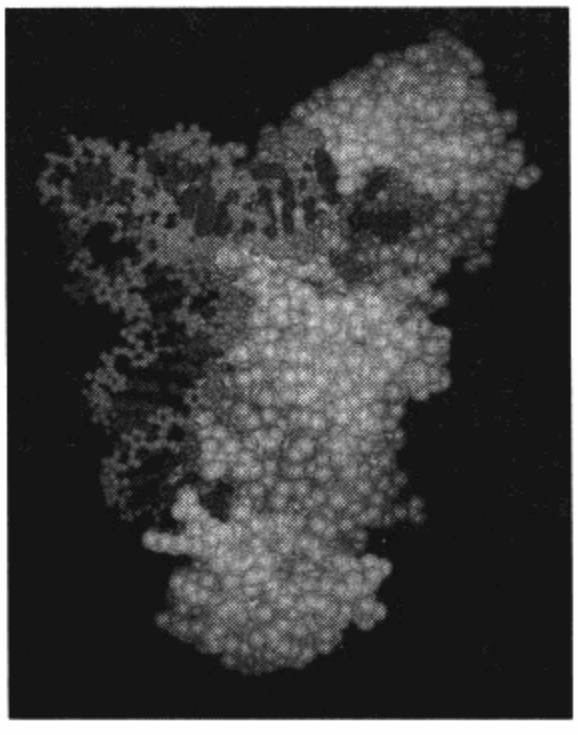


图 6-60　氨酰 tRNA 合成酶对 tRNA 分子的识别

对于该 tRNA 合成酶（$tRNA^{Gln}$）而言，反密码子（底部）和氨基酸接受臂中的特异核苷酸使正确的 tRNA 能够被合成酶识别（蓝色）（经 Tom Steitz 允许后使用）

氨基酸加入到生长多肽链的 C 端

我们已经看到，氨基酸首先偶联到 tRNA 分子上，现在我们再看看将这些氨基酸连接起来形成蛋白质的机制。蛋白质合成的基本反应就是在生长多肽链末端的羰基和即将进入的氨基酸自由氨基之间形成一个肽键。因此，蛋白质是从 N 端逐步合成的。在整个过程中，多肽链生长中的羧基末端通过与 tRNA 分子（氨酰 tRNA）的共价结合，持续激活。这种高能共价连接在每次加成时都受到破坏，但是立即由最新加入的氨基酸上相同的共价连接取代（图 6-61）。通过这种方式，每个加入的氨基酸携带了用于下一个氨基酸加成的活化能，而不是用于它自身加成的能量，图 2-68 列举了“头部生长”类型多聚化的一个例子。

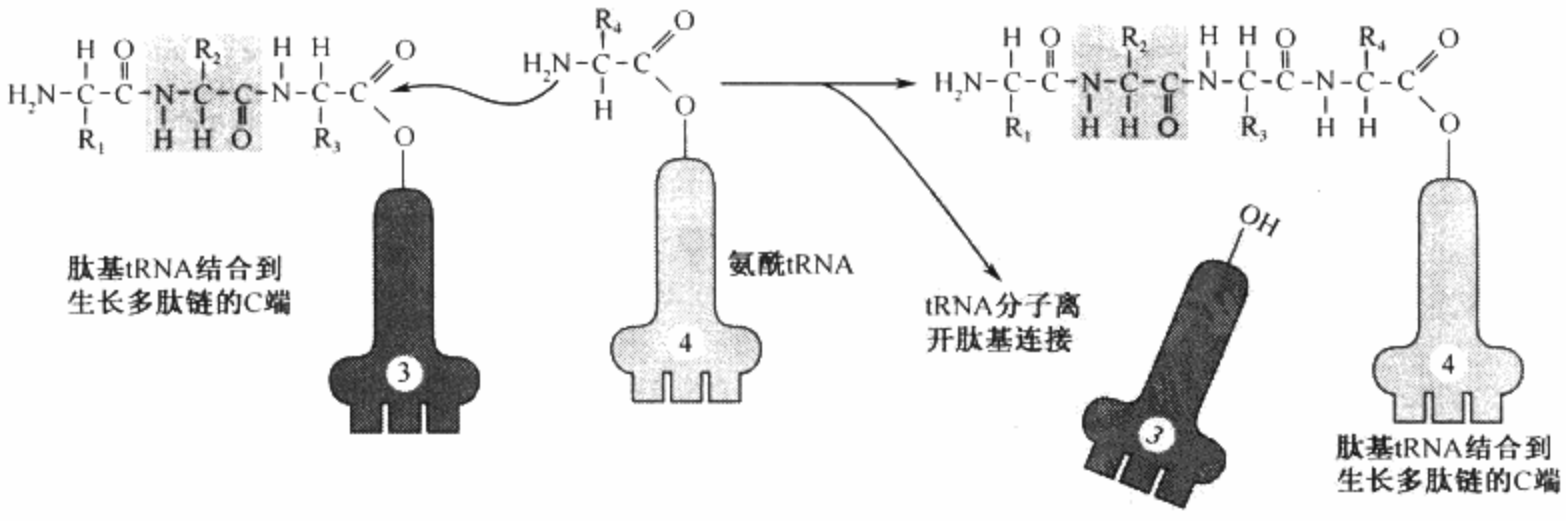


图 6-61　氨基酸整合进入蛋白质

多肽链通过逐步加入氨基酸向其 C 端生长。每个肽键的形成在能量上是有利的，因为 tRNA 分子的共价相连已将生长的 C 端活化。每次加成时，都重新生成活化生长末端的肽酰-tRNA 连接。氨基酸侧链缩写成 R_1、R_2、R_3和 R_4；作为参考位点，肽链中第二个氨基酸所有原子都用灰色阴影表示。该图显示了第 4 个氨基酸加入到生长链中的过程

RNA 信息在核糖体上解译

我们已经看到，蛋白质合成由 mRNA 分子中携带的信息指导，为了维持正确的可读框以及确保精确度（大约每 10 000 个氨基酸发生一次错误），蛋白质的合成在**核糖体**中进行。核糖体是由至少 50 种不同蛋白质（核糖体蛋白）和多种 RNA 分子、**核糖体 RNA（rRNA）**组建起来的一种复杂的催化装置。一个典型的真核细胞在它的细胞质内含有数百万个核糖体（图 6-62）。我们已经看到，真核生物核糖体亚基在细胞核内组装，是将新转录的且经过修饰的 RNA 与细胞质内合成后输入细胞核的核糖体蛋白结合起来。然后两个核糖体亚基输出到细胞质中，在那里行使蛋白质合成的功能。

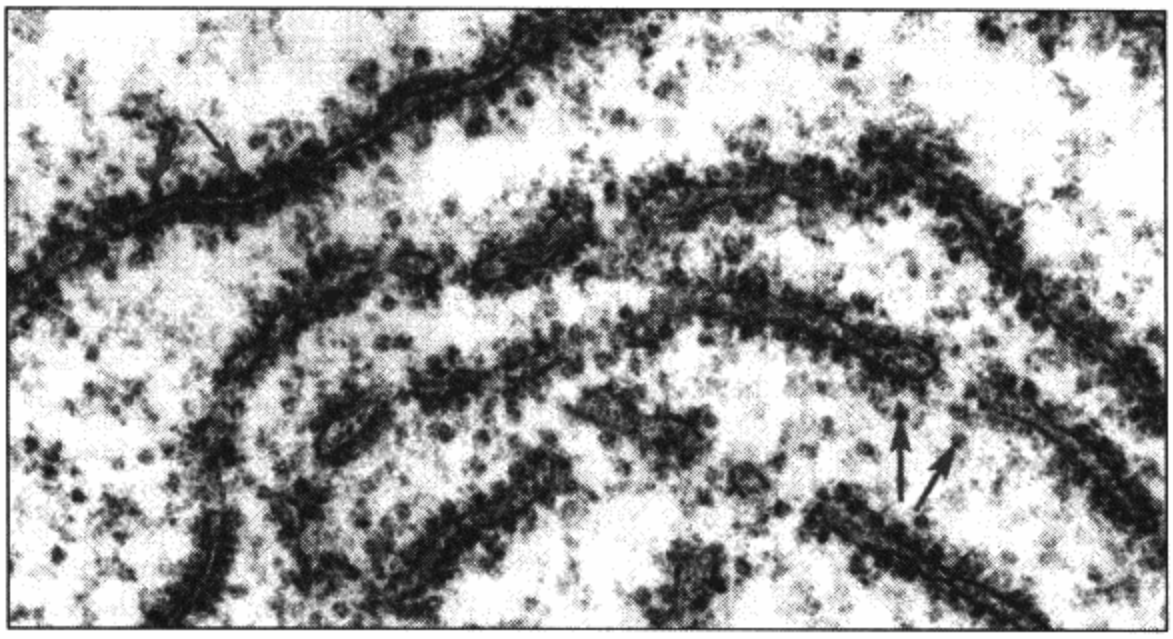


图 6-62　位于真核细胞质中的核糖体

该电镜显微照片显示了小块细胞质区域的截面图。黑点表示核糖体（红色箭头）。它们有些在细胞质内游离存在；其他的附着到内质网上（经 Daniel S. Friend 允许后使用）

真核生物与原核生物的核糖体在结构和功能上都非常相似。二者都由一大一小两个亚基组成，它们装配在一起形成一个质量达几百万道尔顿的完整核糖体（图 6-63）。小亚基为 tRNA 提供一个与 mRNA 密码子精确匹配的框架（图 6-58）；而大亚基催化肽键的形成，将氨基酸连接起来形成一条多肽链（图 6-61）。核糖体的这两个亚基在没有活跃地合成蛋白质时发生解离。它们常常在 mRNA 分子 5′端附近组装起来，起始一个蛋白质的合成。然后 mRNA 被牵引通过核糖体，当 mRNA 的密码子遇到核糖体活性位点时，就利用 tRNA 作为转换器，将每个氨基酸按正确的顺序加到生长多肽链末端，从而将 mRNA 序列翻译成氨基酸序列。当遇到终止密码子时，核糖体就释放已完成的蛋白质，它的两个亚基再次解离，然后这些亚基可以用于在另一条 mRNA 分子上起始另一个蛋白质的合成。

核糖体工作效率非常高：每秒钟真核细胞每个核糖体可加入大约 2 个氨基酸到多肽链上；细菌细胞的核糖体甚至运行得更快，速率大约是每秒 20 个氨基酸。核糖体是怎样设计有效翻译所需的多种协同运动的呢？每个核糖体含有 4 个结合 RNA 的位点：一个用于 mRNA 结合，其他 3 个用于（称为 A 位点、P 位点和 E 位点）tRNA 结合（图 6-64）。tRNA 分子只有当它的反密码子与核糖体上结合的 mRNA 分子的互补密码子形成碱基配对时，才能牢牢地维

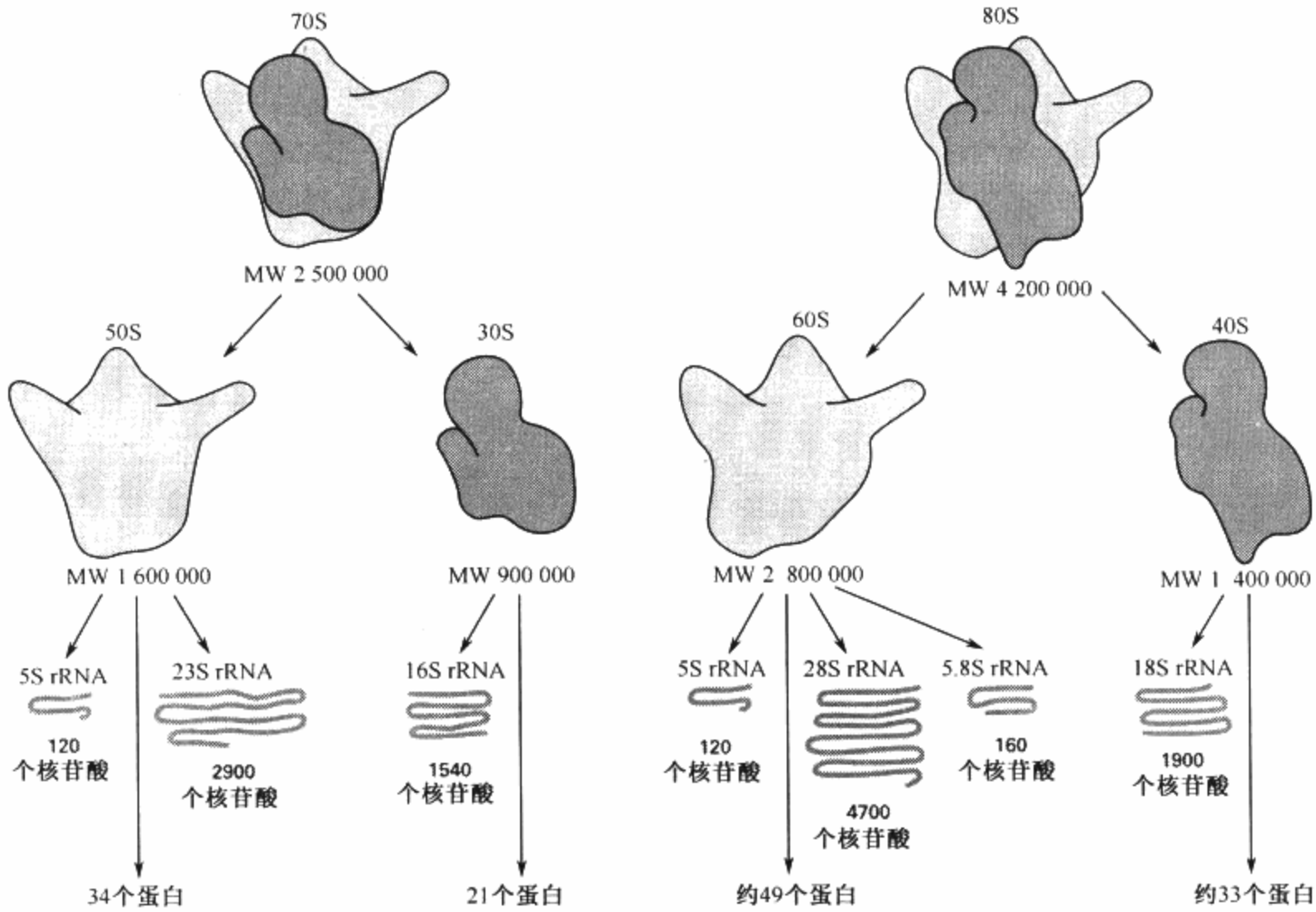


图 6-63　原核生物与真核生物的核糖体结构比较

核糖体组分一般根据它们的“S 值”命名，S 值是指在超速离心时的沉降速率。尽管它们的 rRNA 和蛋白质成分的数量与大小不同，但原核生物和真核生物核糖体具有几乎相同的结构，发挥着相似的功能。虽然真核生物核糖体的 18S rRNA 和 28S rRNA 含有许多在细菌核糖体中不存在的其他核苷酸，但这些核苷酸是作为多种插入而存在的，它们形成额外的区域，并使每个 rRNA 基本结构大部分保持不变

系在 A 和 P 位点。这两个位点距离足够近，以便迫使它们的 2 个 tRNA 分子能与 mRNA 分子上的相邻密码子形成碱基配对。核糖体的这个特征维持了 mRNA 上正确的可读框。

一旦已经起始了蛋白质合成，通过含有 3 个主要步骤的循环反应，每个新氨基酸就会加入到正在延伸的肽链上。我们对肽链延伸过程的描述开始于某个位置，在这里已经有一些氨基酸连接起来，并在核糖体的 P 位点有一个 tRNA 分子，与生长的多肽链末端共价相连（图 6-65）。在第一步，携带多肽链下一个氨基酸的 tRNA 通过与位于 A 位点处的 mRNA 的密码子形成碱基对，结合到核糖体的 A 位点上，这样 A 位点和 P 位点就结合有相邻 tRNA。在第二步，多肽链的羧基末端从 P 位点的 tRNA 上释放（通过打断 tRNA 和它的氨基酸之间的高能键），并与 P 位点的 tRNA 上结合的氨基酸的自由氨基相结合，形成一个新的肽链。蛋白质合成的这个中心反应由核糖体拥有的一种肽键转移酶催化活性进行催化。这个反应伴随核糖体中许多构象变化，由它们将 2 个 tRNA 转移到大亚基的 E 位点和 P 位点。在第三步，另外一系列的构象变化使 mRNA 移动穿过核糖体恰好 3 个核苷酸，并重新复位核糖体，这样它就准备好接受下一个氨酰-tRNA。然后，与一个新加入的氨酰-tRNA 一起重复第一步，依此类推。

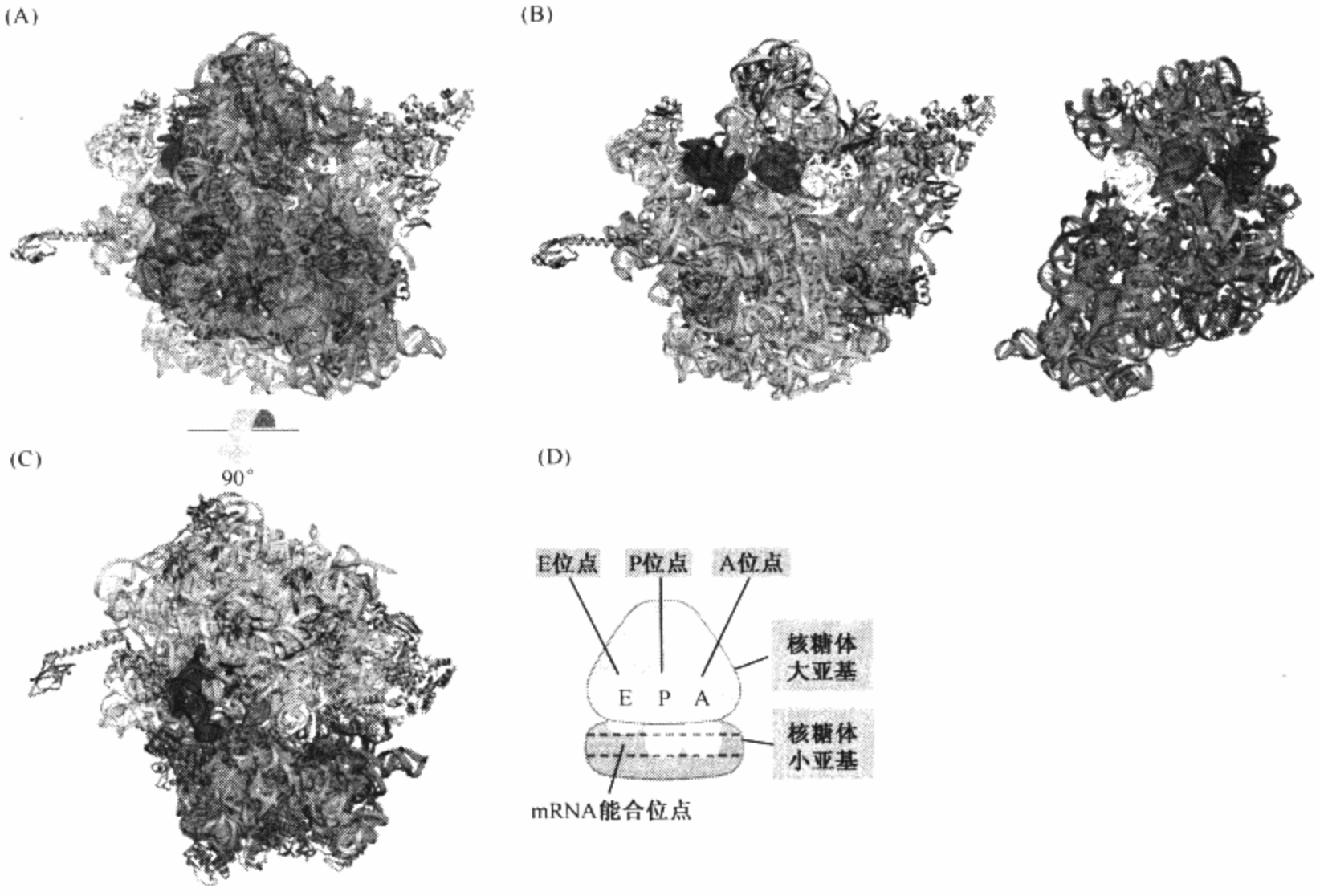


图 6-64　核糖体中的 RNA 结合位点

每个核糖体有 3 个 tRNA 结合位点：A、P 和 E 位点（分别为氨酰-tRNA、肽酰-tRNA 和出口的简写）和 1 个 mRNA 结合位点。（A）细菌核糖体的结构，前面的是小亚基（深绿色），后面的是大亚基（浅绿色）。rRNA 和核糖体蛋白都有显示。所显示的 tRNA 结合在 E 位点（红色）、P 位点（橘黄色）和 A 位点（黄色）。虽然全部 3 个位点在这里都显示被占据着，但是据认为在蛋白质合成过程中，这些位点最多只有 2 个在任何时候都有 tRNA 存在（图 6-65）。（B）核糖体大亚基（左边）和小亚基（右边）的结构在核糖体中的排列就好比将（A）中的核糖体像一本书似地打开了。（C）从上部观察的（A）中核糖体结构。（D）核糖体的高度概括图示［与（C）中的方向一致］，它在后面的图中也将会用到［（A）、（B）和（C）摘自 M. M. Yusupov et al., *Science* 292：883～896，2001，经 Albion Bausom and Harry Noller 允许后使用］

每次一个氨基酸加入到多肽链就重复这个三步组成的循环，而多肽链也从氨基端向羧基端生长，直到遇上一个终止密码子。

延伸因子推动翻译向前进行

图 6-65 中概括的多肽链延伸的基本循环具有另一个特征，使得翻译特别有效，特别精确。每一轮有 2 个延伸因子（EF-Tu 和 EF-G）进入并离开核糖体，在此过程中它们每个都将 GTP 水解成 GDP，并发生构象变化。在有些条件下，可以使核糖体在没有延伸因子和 GTP 水解的帮助下进行蛋白质合成，但是这个合成过程很慢，效率低而且不精确。如果将延伸因子的构象变化与核糖体不同构型状态转变之间偶联起来，就会大大加快这个过程。虽然有关核糖体中这些构型变化的细节并不清楚，但是有些可能涉及类似 RNA 剪接过程中的 RNA 重排（图 6-30）。延伸因子结合、GTP 水解和解离的这个循环确保了构型变化朝“前进”的方向进行，因此翻译能够有效地进行（图 6-66）。

EF-Tu 除了帮助翻译向前进行之外，还认为它通过监控带电荷的 tRNA 和密码子之间的

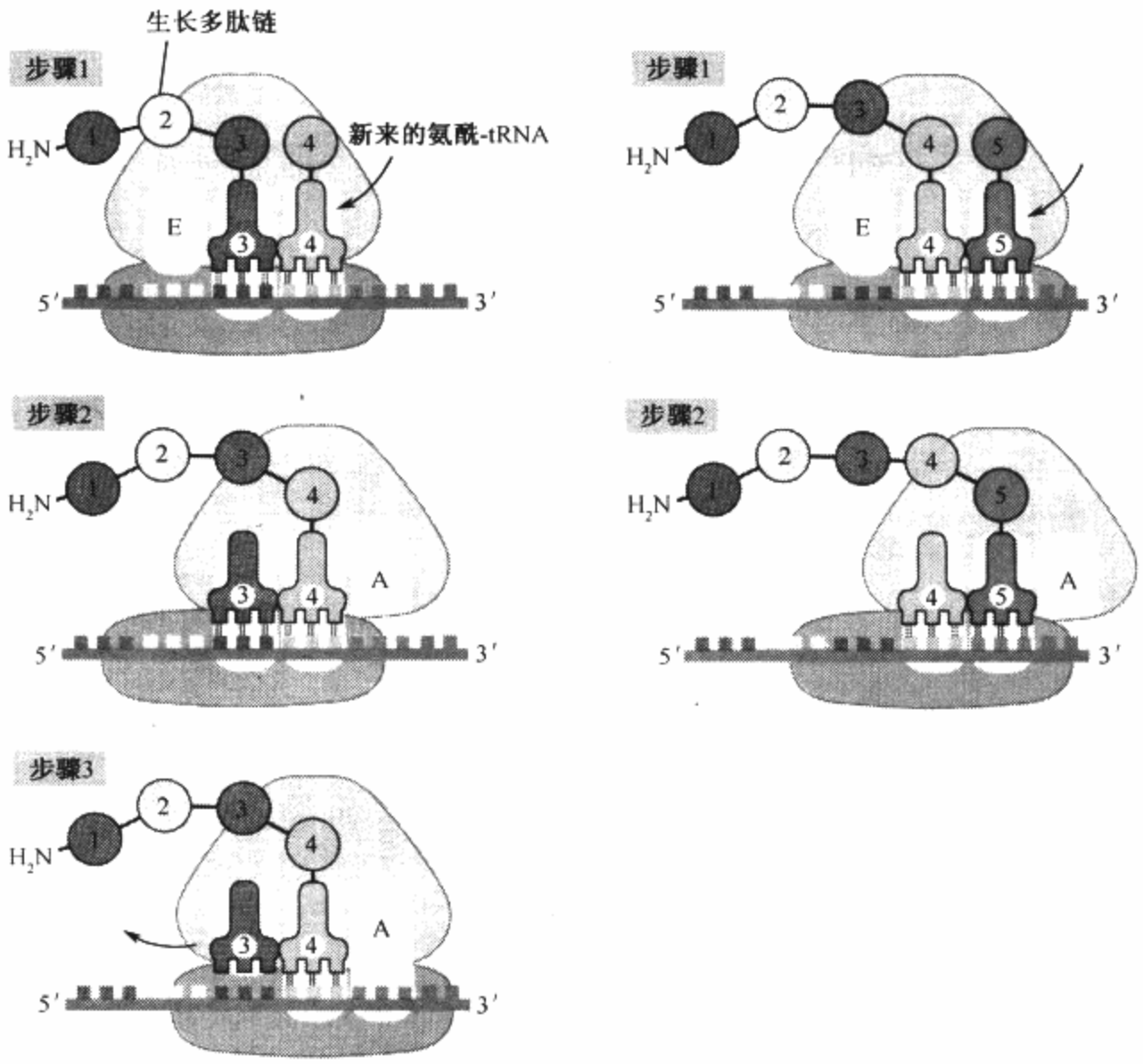


图 6-65　mRNA 分子的翻译

每个加入到多肽链生长末端的氨基酸，由携带氨基酸的 tRNA 分子中的反密码子与 mRNA 链中的下一个密码子之间互补碱基配对进行选择。由于细胞内许多种 tRNA 分子中只有一种能与每个密码子形成碱基对，所以密码子决定了将要加入到生长多肽链中的特定氨基酸。图中所示的三步反应循环在蛋白质合成过程中一次又一次重复。第一步，一个氨酰-tRNA 分子结合到核糖体上空着的 A 位点；第二步，形成一个新的肽键；第三步，mRNA 沿小亚基链移动 3 个核苷酸，同时释放出用过的 tRNA 分子并“重置”核糖体，这样下一个氨酰 tRNA 分子才能结合。虽然该图显示小亚基相对于大亚基有较大的位移，但在翻译过程中核糖体内实际发生的构象变化更加细微。有可能它们涉及每个亚基内部微小的重排以及两个亚基之间多次微小的位移。如图所示，mRNA 以 5′→3′方向翻译，首先合成的是蛋白质的 N 端，每一轮循环向多肽链的 C 端增加一个氨基酸。生长多肽链附着到 tRNA 上的位置不发生改变：它总是连接到位于核糖体大亚基 P 位点的 tRNA 上

最初的相互作用，提高了翻译的精确性（图 6-66）。带电的 tRNA 进入结合有 EF-Tu 的核糖体中，此时 EF-Tu 处于 GTP 结合形式。虽然这个结合的延伸因子允许进行密码子-反密码子配对，但是它阻止氨基酸整合到生长的多肽链上。起初的密码子识别，引起延伸因子水解它结合的 GTP（水解为 GDP 和无机磷酸），于是延伸因子很快从核糖体上解离，不再带有 tRNA，从而使蛋白质合成能向前进行。延伸因子导致在密码子-反密码子碱基配对和多肽链延伸之间出现两次短暂的停顿，这两次停顿选择性地让错误结合的 tRNA 在不可逆转的链延伸步骤之前从混合体出来。第一次停顿是因为需要时间用来水解 GTP。对于正确配对的密码子-反密码子，由 EF-Tu 进行的 GTP 水解速率比错误配对的要快；因而错误结合的 tRNA 分子有更长时间从核糖体上解离。换句话说，GTP 水解选择性地捕获了正确结合的 tRNA。另一

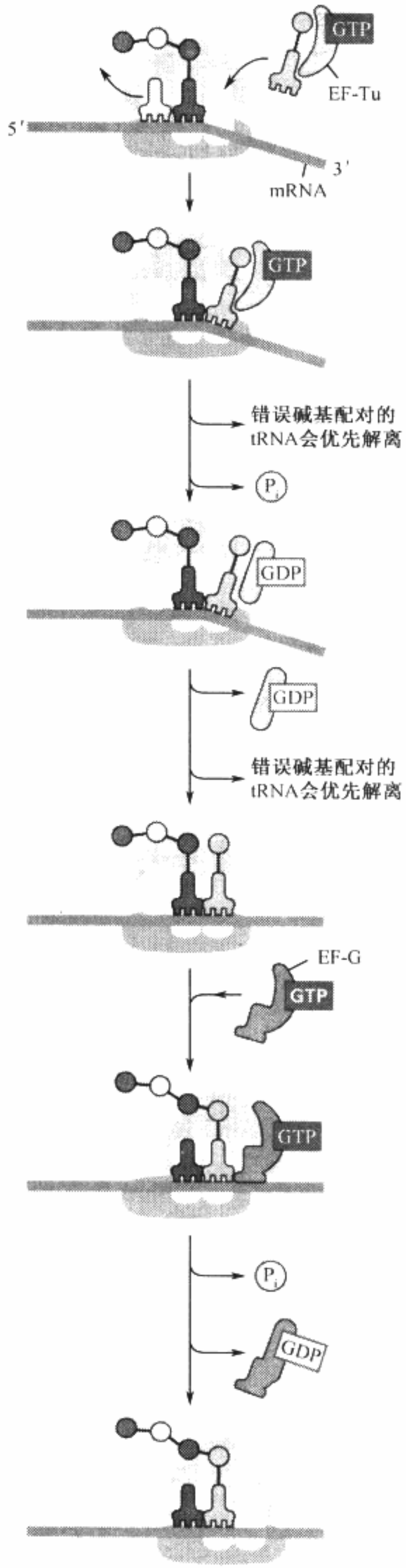


图 6-66　翻译循环详图

图 6-65 中提供的翻译要点还具有其他特征，包括延伸因子以及一种提高翻译精确性机制的参与。在起始结合事件中（最上面的图），一个牢固结合到 EF-Tu 上的氨酰 tRNA 分子与小亚基 A 位点的密码子形成短暂的碱基配对。在此过程中（第 2 个图），tRNA 占据核糖体上的杂合结合位点。密码子-反密码子配对引发 EF-Tu 水解 GTP，从而引起它从氨酰-tRNA 上解离，现在这个氨酰 tRNA 进入 A 位点（第 4 个图），并能参与链延伸反应。因而在蛋白质合成机制中，氨酰-tRNA 的结合和它可用于蛋白质合成之间存在延迟。如文中所述，这个延迟增加了翻译的精确性。在后面的步骤中，以 GTP 结合形式存在的延伸因子 EF-G 进入核糖体，并结合到核糖体大亚基上的 A 位点内或附近，从而使 2 个结合的 tRNA 加速移动进入 A/P 和 P/E 杂合状态。与核糖体的接触刺激了 EF-G 的 GTPase 活性，当它从 GTP 结合形式转变为 GDP 结合形式时，引起 EF-G 中显著的构象变化。这种变化将与 A/P 杂合状态结合的 tRNA 移动到 P 位点并使翻译循环向前推进一个密码子。在每一轮翻译延伸中，tRNA 分子以一系列复杂的旋转移动穿过核糖体，在此过程中它们短暂占据多个“杂合”结合状态。在一种状态下，tRNA 同时结合小亚基的 A 位点和大亚基的 P 位点；而在另一状态下，tRNA 结合小亚基的 P 位点和大亚基的 E 位点。一轮循环中，一个 tRNA 分子据认为占据了 6 种不同的位点，起始结合位点（称为 A/T 杂合位点）、A/A 位点、A/P 杂合状态、P/P 位点、P/E 杂合状态和 E 位点。还认为每个 tRNA 都像单向齿轮一样滚动经过这些位点，在每个发生变化的局部位置沿它的长轴发生旋转。EF-Tu 和 EF-G 是给细菌延伸因子的命名；真核生物中它们分别称为 EF-1 和 EF-2。图 3-74 说明了 GTP 水解引起 EF-Tu 三维结构发生显著的变化。对于所形成每个肽键，EF-Tu 和 EF-G 分子都以它们无活性的 GDP 结合形式释放；为了能再次利用，这些蛋白质必须用 GTP 交换 GDP。对 EF-Tu 而言，这种交换由称作 GTP 交换因子的一大类蛋白质中一个特殊成员完成

次停顿发生在 EF-Tu 解离和 tRNA 完全进入核糖体 A 位点之间。虽然认为这个停顿对于正确和错误结合的 tRNA 都是一样的，可是错误 tRNA 分子形成的密码子-反密码子氢键数目比正确配对的要少，所以在此期间更有可能解离。延伸因子造成的这两次停顿引起大部分错误

结合的 tRNA 分子（以及很多正确结合的分子）离开核糖体，不再用于蛋白质合成，这个两步机制负责核糖体合成蛋白质时 99.99%的精确性。

最近的发现表明，EF-Tu 可能在提高反应的总体精确性时也有其他的作用。本章前面，我们讨论了氨酰合成酶在将氨基酸与 tRNA 精确匹配过程中的作用。当 GTP 结合形式的 EF-Tu护送氨酰-tRNA 到核糖体上时（图 6-66），它明显对氨基酸与 tRNA 之间正确的对应关系进行了双重核查，并将不匹配的扔掉。这个过程究竟是如何进行的还不太了解，但是它可能涉及 EF-Tu 和氨酰-tRNA 之间的全部结合能。根据这个观点，对 EF-Tu 而言，正确匹配的亲和力的范围很窄，而错误的匹配则要么太强要么太弱。因此，EF-Tu 似乎在许多不同的氨基酸-tRNA 组合中进行识别，虽然比较粗略，可是选择性地只让正确的组合进入核糖体。

核糖体是一种核酶

核糖体是由 2/3 的 RNA 和 1/3 的蛋白质组成的一个巨大而复杂的结构。2000 年得到它的大小亚基整体三维结构是现代结构生物学取得的主要的成就。得到的结构有力地证实了之前表明 RNA（而不是蛋白质）对核糖体的整体结构起决定作用，RNA 能将 tRNA 定位到 mRNA 上，RNA 能催化形成共价肽键的证据。例如，RNA 折叠成高度紧密的、精确的三维结构，构成了核糖体的核心，从而决定了它的整体形状（图 6-67）。

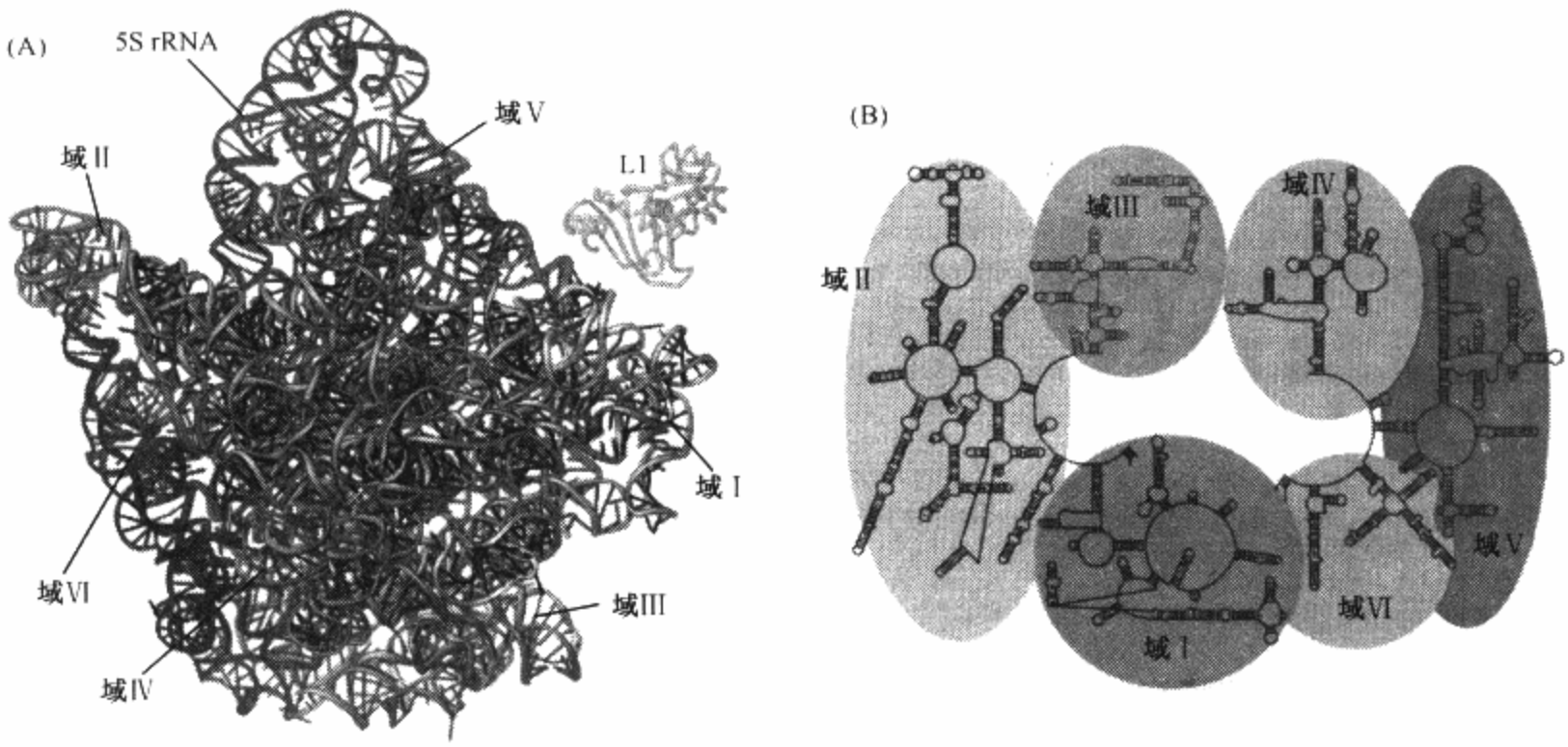


图 6-67　利用 X 射线晶体学得到的细菌核糖体大亚基中的 rRNA 结构

（A）rRNA（5S 和 23S）在核糖体大亚基中存在时的三维结构。还显示了核糖体的其中一个蛋白质亚基（L1）作为参照点，因为它在核糖体上形成一个特征性的突出。（B）23S rRNA 的二级结构简图，显示了广泛的碱基配对网络。整个结构被分成 6 个结构“域”，它们的颜色对应于（A）图三维结构中的颜色。二级结构图高度概括以便能够在二维水平上尽量多地代表真实结构。为了做到这一点，引入 RNA 中多个不连续区域，虽然实际上 23S rRNA 只是单独的一个 RNA 分子。例如，区域Ⅲ的碱基与区域Ⅳ的碱基是连续的，尽管图中出现了一个缺口（摘自 N. Ban et al., *Science* 289：905～920，2000）

与 rRNA 位于中心位置形成鲜明对比的是，核糖体蛋白通常位于表面，填补折叠 RNA 的缺口和缝隙（图 6-68）。这些蛋白质中有的含有球形结构域，略微伸进 RNA 核心中的洞口

内（图 6-69）。核糖体蛋白的主要功能似乎是稳定 RNA 核心，同时允许发生该 RNA 催化有效的蛋白质合成所必需的 rRNA 构象变化。

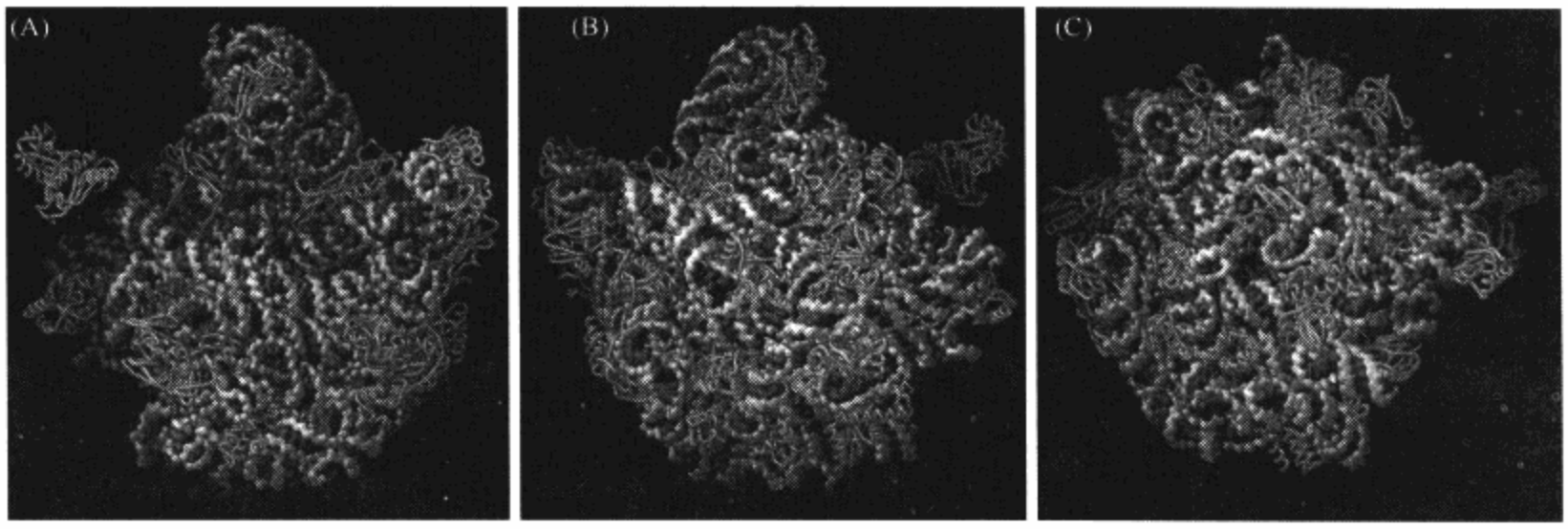


图 6-68　细菌核糖体大亚基蛋白质成分的定位

rRNA（5S 和 23S）描绘成灰色，大亚基蛋白（31 种中的 27 种）描绘成金色。为了方便，所描绘的蛋白质结构只有多肽骨架。（A）与小亚基交界面的视图，与图 6-64（B）中所显示的图一样。（B）大亚基后部视图，将（A）图绕垂直轴旋转 180°后得到。（C）大亚基底部视图，显示了该结构中央的肽出口通道（摘自 N. Ban et al., *Science* 289：905～920，2000 © AAAS）

不仅核糖体上供 tRNA 结合的 3 个位点（A、P 和 E 位点）主要由核糖体 RNA 形成，而且很明显催化肽键形成的位点也由 23S RNA 形成，离该位点最近的氨基酸至少在 1.8nm 以外。在很多方面肽酰转移酶的这个以 RNA 为基础的位点与有些蛋白质中发现的位点相似；它是一个有高度构造的口袋，精确地对两种反应物定向（生长的肽链和一个氨酰-tRNA），并且它提供了一个功能基因，作为一种通用的酸-碱基催化剂。本例中显然是腺嘌呤的氮环，而不是氨基酸侧链，如组氨酸残基（图 6-70）。RNA 分子作为这样一种催化剂发挥作用的能力最初令人奇怪，因为人们认为 RNA 缺少既能接受质子，又能供出质子的合适的化学基团。虽然腺嘌呤氮环的 pK 值常常是 3.5 左右，可是 23S rRNA 活性位点的三维结构及电荷分布迫使这个看起来很关键的腺嘌呤的 pK 值处于中性范围，因而产生了酶的催化活性。

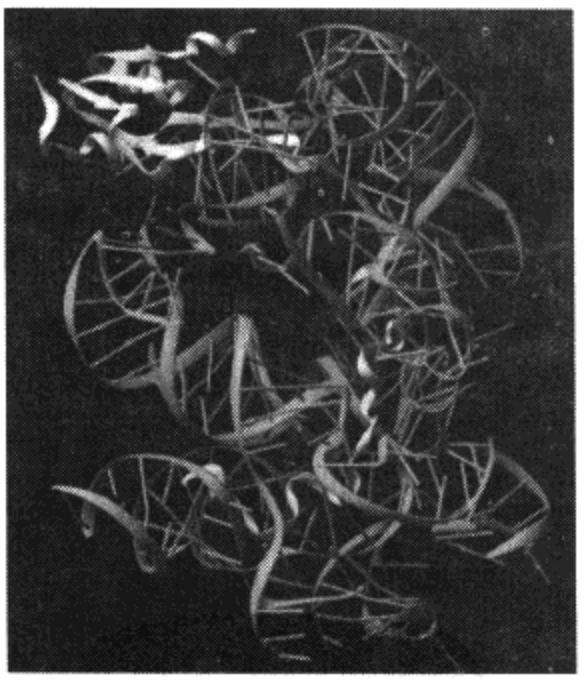


图 6-69　细菌核糖体大亚基中 L15 蛋白的结构

该蛋白质的球状结构域位于核糖体表面，一个伸展区深深插入到核糖体的 RNA 核心中。图中黄色的是 L15 蛋白，红色的是核糖体 RNA 核心的一部分（经 D. Klein，P. B. Moore 和 T. A. Steitz 允许后使用）

具有催化能力的 RNA 分子称为**核酶**。本章前面我们看到其他核酶在 RNA 剪接反应中是如何起作用的（图 6-36）。在本章的最后一部分，我们探讨最新发现的作为催化剂发挥作用的 RNA 分子催化许多不同反应的能力，这对于早期活细胞的进化可能意味着什么。这里我们只需要注意到，有很好的理由去怀疑 RNA 而不是蛋白质分子作为活细胞最早的催化剂发挥功能。如果是这样，核糖体与

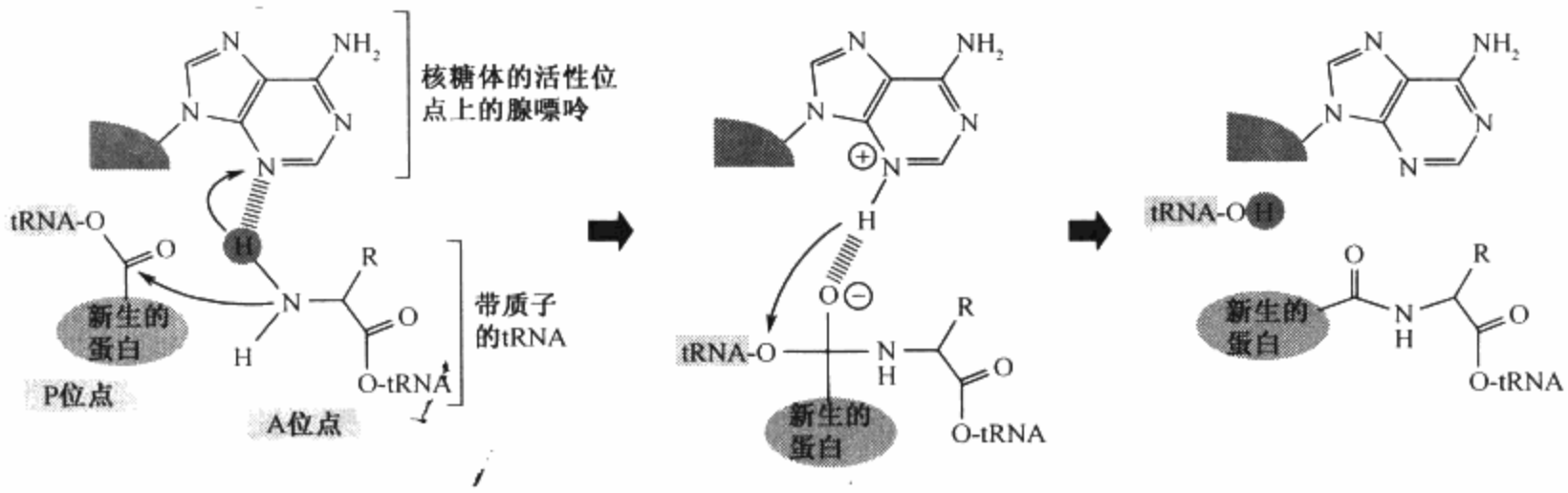


图 6-70　存在于核糖体大亚基中的肽酰转移酶活性可能的反应机制

整个反应由 23S rRNA 中的活性位点催化。在所提出的这个机制第一步中，活性位点腺嘌呤的 N3 抽走核糖体 A 位点 tRNA 上附着氨基酸的一个质子，使其氨基酸能够攻击多肽链生长末端的羟基。下一步这个质子化腺嘌呤贡献它的氢给连接到肽酰 tRNA 上的氧，引起这个 tRNA 从肽链释放。这样就生成一条比起始反应物多一个氨基酸的多肽链，然后整个反应循环将以下一个氨酰-tRNA 进入 A 位点重复（改编自 P. Nisson et al.，*Science* 289：920～930，2000）

它的 RNA 核心一起，可以看成是生命历史早期的遗迹，那时，蛋白质合成在那些几乎全由核酶催化的细胞中进化。

mRNA 中的核苷酸序列提供了从哪里开始合成蛋白质的信号

翻译的起始和终止通过改变上面描述的翻译延伸循环来实现。mRNA 上的蛋白质合成起始的位点很关键，因为它决定了完整信息的可读框。在这个时期无论向前或是向后错位一个核苷酸，都将导致信息中后面的密码子阅读错误，因而就会产生一个序列删改的无功能蛋白质。从另一方面看，起始步骤也极其重要，因为对于大多数基因来说，这是最后一个关节点，细胞能够决定是否翻译 mRNA，合成蛋白质，因而起始的速率决定了蛋白质合成的速率。第 7 章我们将看到细胞利用几种机制调节翻译的起始。

mRNA 的翻译从密码子 AUG 开始，并需要一种特殊的 tRNA 来起始翻译。这个**起始 tRNA** 常常携带有甲硫氨酸（在细菌中是一种修饰形式的甲硫氨酸——甲酰甲硫氨酸），这样所有新合成的蛋白质 N 端第一个也就是蛋白质首先合成的末端氨基酸都是甲硫氨酸，这个甲硫氨酸后来通常由一种特殊的蛋白酶去除。起始 tRNA 核苷酸序列与普通的携带甲硫氨酸的 tRNA 的序列不同。

在真核生物中，起始 tRNA（偶联甲硫氨酸）首先与其他蛋白质，称作**真核起始因子**，或 **eIF**（图 6-71）一起结合到小核糖体亚基上。细胞所有的氨酰-tRNA 中，只有负责甲硫氨酸的起始 tRNA 能够牢固地结合核糖体小亚基，而不需要完整核糖的存在。下一步，核糖体小亚基结合到 mRNA 分子的末端，该 mRNA 分子借助其帽子以及结合的两个起始因子，eIF4E（直接结合在帽子上）和 eIF4G（图 6-40）进行识别。然后核糖体小亚基沿着 mRNA 向前移动（5′→3′），寻找第一个 AUG。这种移动受到其他起始因子的促进，它们像 ATP 推动的解螺旋酶那样发挥作用，从而使小亚基能够扫描通过 RNA 的二级结构。在 90%的 mRNA 中，翻译从小亚基遇到的第一个 AUG 开始。在此位点上起始因子从小亚基上解离，为大亚基和它组装成完整的核糖体开路。现在起始 tRNA 结合到 P 位点，A 位点则空出来。蛋白质合成因而就装备好添加下一个氨酰-tRNA 分子（图 6-71）。

真核生物 mRNA 起始位点附近的核苷酸在上述扫描过程中影响 AUG 识别的效率。如果

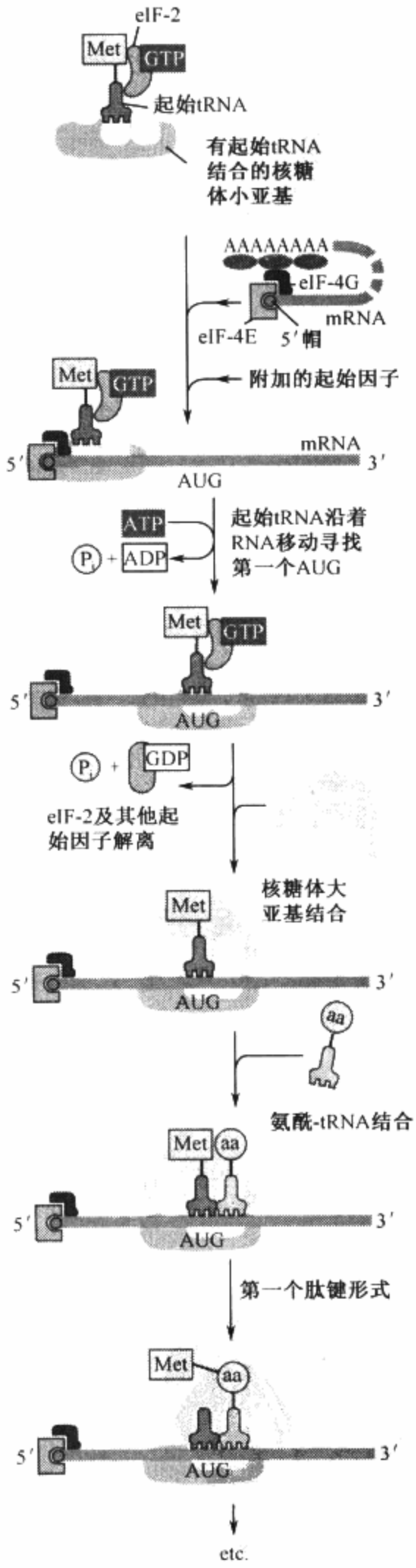


图 6-71　真核生物中蛋白质合成的起始阶段

在该过程所需的许多翻译起始因子中只显示了其中三种。有效的翻译起始也需要 mRNA 的多聚 A 尾，它结合着多聚 A 结合蛋白，这种蛋白质继而与 eIF4G 相互作用。通过这种方式，翻译装置确保 mRNA 两个末端在起始之前是完好的（图 6-40）。虽然图中只显示了一次 GTP 水解事件，但已知就在核糖体大小亚基结合之前有第二个 GTP 发生水解

这个识别位点与共有识别序列差别相当大，核糖体扫描亚基有时会忽略 mRNA 中第一个 AUG 密码子，而跳到第二个或第三个 AUG 密码子上。细胞频繁地利用这种称作“渗漏扫描”的现象，从同一 mRNA 分子合成两种或更多的 N 端不同的蛋白质。例如，它使有些基因能合成 N 端有或没有信号序列的相同蛋白质，这样该蛋白质就可以导向细胞内两个不同的区室。

在细菌中，起始密码子的选择不一样。细菌 mRNA 没有 5′帽子来告诉核糖体从哪儿开始搜寻翻译起始位点。取而代之，每个细菌 mRNA 含有一个特殊的核糖体结合位点（称为 Shine-Dalgano 序列，以其发现者命名），它位于翻译起始 AUG 上游几个核苷酸序列处。这段核苷酸序列的共有序列是 5′-AGGAGGU-3′，它与核糖体小亚基的 16S rRNA 形成碱基配对，从而将起始密码子 AUG 定位到核糖体中。一组翻译起始因子协调了这种相互作用，以及随后进行的核糖体大亚基的装配，形成一个完整的核糖体。

由此可见，与真核生物核糖体不同，细菌核糖体容易直接在位于 mRNA 分子内部的起始密码子上装配，只要核糖体结合位点位于它前面几个核苷酸处。结果，细菌 mRNA 通常都是多顺反子，即它们编码几种不同的蛋白质，每一个都从同一 mRNA 分子翻译而来（图 6-72）。相比之下，一条真核 mRNA 一般只编码一种蛋白质。

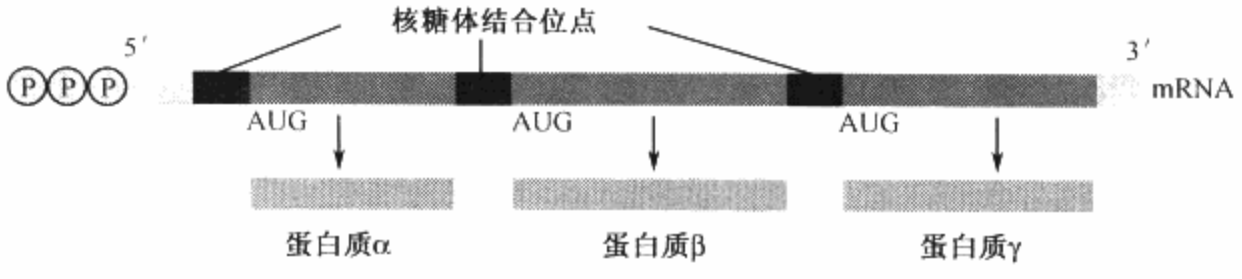


图 6-72　典型的细菌 mRNA 分子结构

与典型真核生物核糖体需要一个加帽的 5′端不同，原核生物核糖体在核糖体结合位点（Shine-Dalgarno 序列）起始翻译，该位点可以位于 mRNA 分子中任何位置。核糖体这个特性使细菌能够从一条 mRNA 分子合成不止一种蛋白质

终止密码子标记翻译的结束

3 个密码子（UAA、UAG 或 UGA）中任何一个的存在，都为蛋白质编码信息提供了结束的信号，称它们为终止密码子（图 6-50）。它们不是由 tRNA 识别，也不指令任何氨基酸，而是标示核糖体停止翻译。称作释放因子的蛋白质结合到任何 A 位点有一个终止密码子的核糖体上。这种结合迫使核糖体中的肽酰转移酶催化一个水分子，而不是一个氨基酸，加成到肽酰-tRNA 上（图 6-73），该反应使生长多肽链的羧基末端从对 tRNA 分子的附着状态中游离出来，由于一般只有这种附着将生长多肽链维系在核糖体上，所以完整的多肽链会立即释放进入细胞质。然后核糖体释放 mRNA，并解离成大小亚基，它们能够在另一个 mRNA 分子上装配起来，开始新一轮蛋白质合成。

释放因子提供了一种显著的分子模拟实例，一类大分子借此与一个化学上不相关分子的形状相似。在这个例子中，释放因子的三维结构（全部由蛋白质组成）与 tRNA 分子的结构和电荷分布之间拥有不可思议的相似性（图 6-74）。这种在形状和电荷上的模拟使释放因子能够进入核糖体 A 位点并引起翻译终止。

在翻译过程中，新生的多肽链要移动通过核糖体大亚基内一条巨大且充满水的隧道（约 10nm×1.5nm）[图6-68(C)]。该隧道的四壁主要由 23S rRNA 组成，它就像是较小的疏水表面嵌入到较大亲水表面形成的镶嵌物。由于这种结构不与任何肽结构互补，所以就提供了一种“特氟隆”被膜，多肽链可以容易地从中间滑动。隧道的大小也表明，新生多肽链在通过核糖体的时候，大部分发生了结构破坏，尽管在离开核糖体隧道之前也能形成该蛋白质的许多 α 螺旋区域。当新合成的蛋白质离开核糖体时，必须折叠成能为细胞所利用的三维结构，本章后面我们再讨论这种折叠是如何进行的，但是，首先我们回顾一下翻译过程本身的其他几个方面。

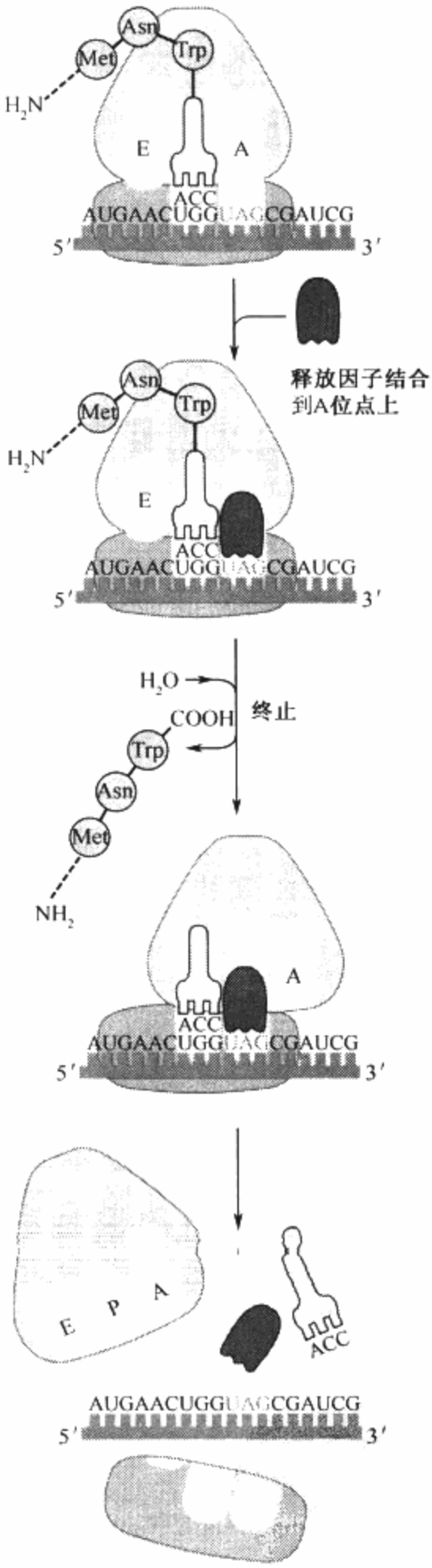


图 6-73　蛋白质合成的最后阶段

一种释放因子结合具有一个终止密码子的 A 位点时终止了翻译。完整的多肽链释放，并且核糖体在核糖体循环因子（未显示）发挥作用以后，分解成它的两个单独亚基

蛋白质在多聚体核糖体上合成

大多数蛋白质的合成需要 20s 到几分钟不等的时间。但是即便在这么短的时间内，每个正在翻译的 mRNA 分子上常常发生多次起始，一旦前面的核糖体已翻译出足够多的核苷酸序列，腾出了位置，mRNA 的 5′端就串上一个新的核糖体，因此常常发现正在翻译的 mRNA 分子处于多聚核糖体的形式（也成为多核糖体），它们是由多个核糖体沿单个 mRNA 分子间隔 80 个核苷酸排列组成的巨大组装物（图 6-75）。这些多重起始过程意味着在给定的时间内，比起每轮合成在下一轮开始之前必须完成来讲，能合成出更多的蛋白质分子。

图 6-74　人类翻译释放因子（eRF1）的结构及其与 tRNA 分子的相似性

左边的是蛋白质，右边的是 tRNA（摘自 H. Sang et al., *Cell* 100：311～321. 2000 © Elsevier）

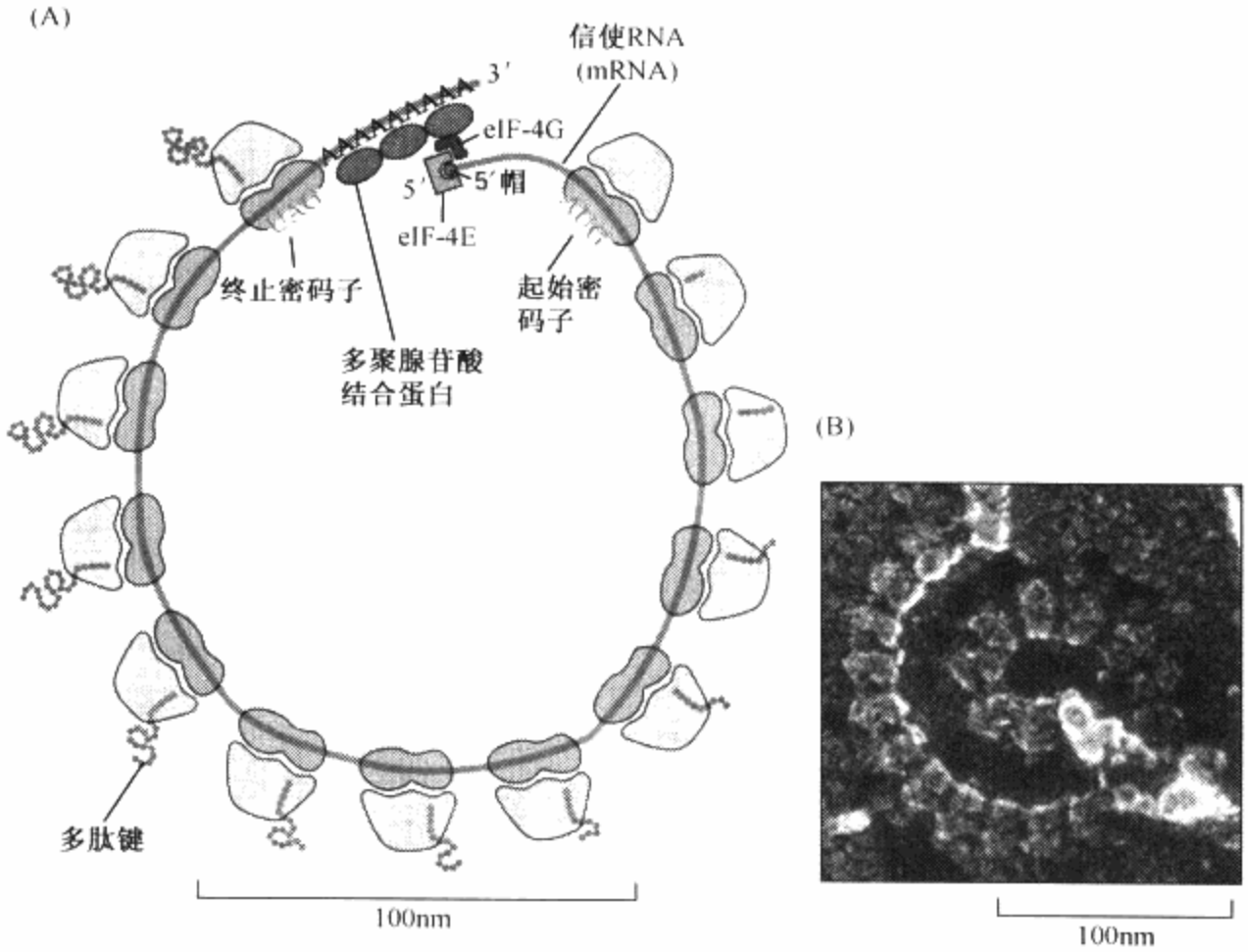


图 6-75　多聚核糖体

（A）显示一系列核糖体怎样同时翻译同一真核 mRNA 分子的简图。（B）来自真核细胞多聚核糖体的电镜显微照片[(B)经 John Heuser 允许后使用]

细菌和真核生物都利用多核糖体合成蛋白质，并且两者都运用了其他的策略，以便使蛋白质合成速度更快。由于细菌的 mRNA 不需要加工，并且还在合成时就能为核糖体所利用，因此，甚至于细菌 mRNA 转录还未完成时，核糖体就能附着到它的自由末端，从而紧跟着 RNA 聚合酶沿 DNA 移动。我们已经看到，真核生物中，mRNA 的 5′端和 3′端相互作用[图 6-40和图6-75(A)]；因此，一旦核糖体解聚，它们两个亚基就位于最佳位置，重新起始同一 mRNA 分子上的翻译。

质量控制机制在翻译的许多阶段发挥作用

核糖体进行的翻译可以看成是一对相反的限制要求（精确性与速率）的折中。例如，我们已经看到，翻译的精确性（每连接 10^4 氨基酸发生 1 次错误）要求每次向生长多肽链加入一个新的氨基酸时发生一次停顿，从而细菌整体翻译速度是每秒整合 20 个氨基酸。核糖体小亚基存在某种特殊改变的细菌突变株，将 mRNA 翻译成蛋白质时显著高于这个精确率；然而，在这些突变体中蛋白质合成速度如此低，以至于细菌很少能够存活下来。

我们还看到，蛋白质合成要实现所观察到的精确率就必须消耗大量的自由能；之所以如此是因为，正如第 2 章讨论的那样，细胞内有序度的任何一点增加都需要付出代价。在大多数细胞内，蛋白质合成比其他任何过程消耗的能量都要多。每生成一个新的肽键至少要断裂 4 个高能磷酸键：有 2 个在连接 tRNA 分子和 1 个氨基酸（图 6-56）时消耗，另外 2 个用于在合成过程本身，推动核糖体上发生的反应循环中的步骤（图 6-66）。此外，每次 tRNA 合成酶将错误连接的氨基酸水解掉时也消耗额外的能量（图 6-59），并且每次错误的 tRNA 进入核糖体时会引起 GTP 水解并被排除时也消耗能量（图 6-66）。为了有效地发挥作用，这些校对机制也必须消除很大一部分的正确相互作用；因为如此，它们甚至比看起来更加消耗能量。

其他质量控制机制确保了真核 mRNA 分子在核糖体对它进行翻译之前是完整的。翻译已经断裂或部分加工的 mRNA 将对细胞有害，因为将会产生截断的或者在其他情况下会生成畸形的蛋白质。我们已经在真核生物中看到，mRNA 合成不仅涉及转录，也涉及一系列复杂的 mRNA 加工步骤；这些发生在细胞核内，与核糖体隔离，并且只有等到加工完成以后，才将 mRNA 转运到细胞质进行翻译（图 6-40）。然而一个完好的 mRNA 分子如果离开细胞核，就会在细胞质中断裂。为了避免翻译这种断裂的 mRNA 分子，在翻译开始前，5′帽子和多聚 A 尾都由翻译起始装置识别（图 6-71 和图 6-75）。

细菌以一种完全不同的方式解决不完整 mRNA 的问题。不仅在细菌 mRNA 的 3′端存在信号，而且，正如我们已看到的，翻译常常也在转录完成之前就开始进行。当细菌核糖体翻译到不完整的 mRNA 末端时，一种特殊 mRNA（称作 tmRNA）进入核糖体的 A 位点，并翻译自身，这样就在截短的蛋白质 C 端增加了一个特殊的 11 氨基酸标签。从而为蛋白酶提供信号，表示该蛋白质将被降解掉（图 6-76）。

标准遗传密码中存在较小的变动

第 1 章讨论过，遗传密码（图 6-50）适用于生命界 3 个主要分支的全部物种。从而提供了重要证据，表明地球上的所有生命存在共同的祖先。尽管稀少，但这套密码也存在例外的情况，这一部分我们讨论其中的一些，如白色念珠菌（*Candida abicans*），最广泛存在的人类真菌病原体，将密码子 CUG 翻译成丝氨酸，然而几乎所有其他生物都将它翻译成亮氨酸。

线粒体（它拥有自己的基因组并编码自身转录装置的大部分组分）也表现出与标准密码子的一些偏离。例如，在哺乳动物线粒体中，AUA 翻译成甲硫酸，而在细胞质中却翻译成异亮氨酸（表 14-3）。

上面所讨论的遗传密码的这种偏离已根深蒂固于它所出现的生物体或细胞器中。许多细胞中也出现另一种不同的变化，有时称为翻译再编码。在这种情况下，mRNA 中存在的其他核苷酸序列信息可以在 mRNA 分子中特定位点改变遗传密码的意义。标准密码使细胞能够仅仅利用 20 种氨基酸来生产蛋白质。然而，细菌、古菌和真核生物拥有第 21 种氨基酸，通过

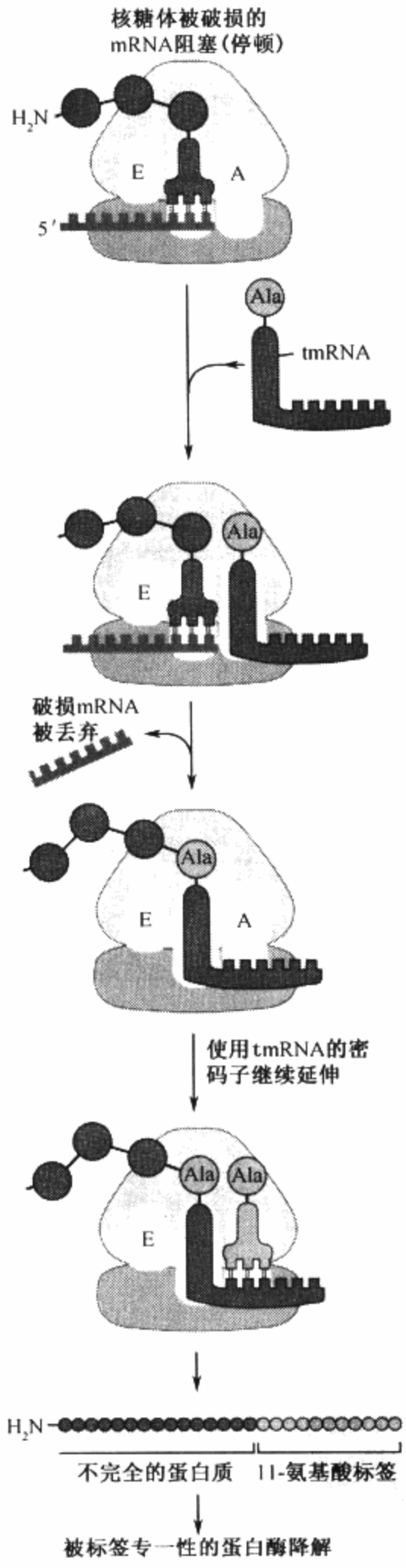


图 6-76　解救停顿在不完整 mRNA 分子上的细菌核糖体

图中显示的 tmRNA 是一段长为 363 个核苷酸的 RNA，既具有 tRNA 功能又具有 mRNA 功能，并由此得以命名。它携带一个丙氨酸，并能进入停顿核糖体空出的 A 位点，将此丙氨酸加到多肽链末端，除了没有密码子指导它以外，其他方面都模拟了 tRNA。然后核糖体从 tmRNA 翻译 10 个密码子，形成了蛋白质上一个 11 氨基酸标签。这个标签由随后降解整个蛋白质的蛋白酶识别

翻译再编码用来直接整合到生长多肽链中。硒半胱氨酸对于许多蛋白酶有效发挥功能很重要，它含有一个硒原子，取代了半胱氨酸中的硫原子。硒代半胱氨酸来自附着在特殊 tRNA 分子上的丝氨酸，该 tRNA 与通常用于代表翻译停止的密码子 UGA 配对。在 UGA 密码子上有硒代半胱氨酸插入的蛋白质的 mRNA 携带了其他核苷酸序列，从而引起再编码事件的发生（图 6-77）。

另一种再编码形式是翻译移码。反转录病毒，一大类真核生物病毒，通常利用了这种再

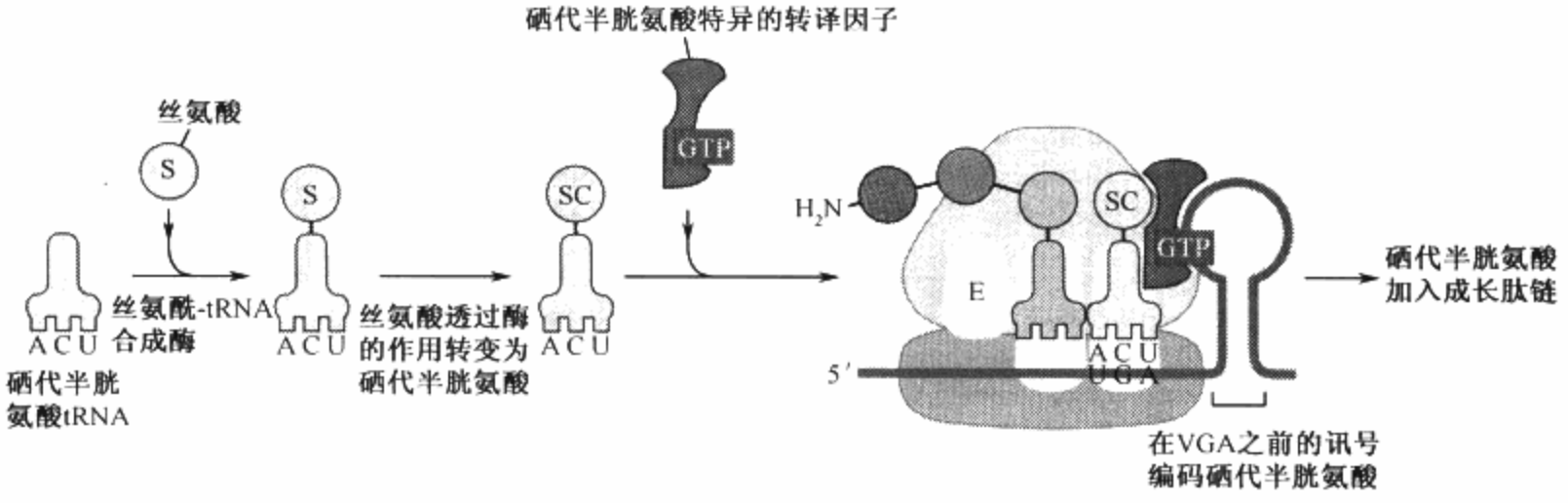


图 6-77　硒代半胱氨酸整合到生长多肽链中

一种特殊的 tRNA 和丝氨酸的连接由正常的丝氨酰-tRNA 合成酶负责，然后这个丝氨酸通过酶的作用转变为硒代半胱氨酸。mRNA 中一种特殊的 RNA 结构（具有特定核苷酸序列的茎环结构）表示该硒代半胱氨酸将要插入到相邻的 UGA 密码子。如图所示，这个过程需要一种硒代半胱氨酸特异的翻译因子参与

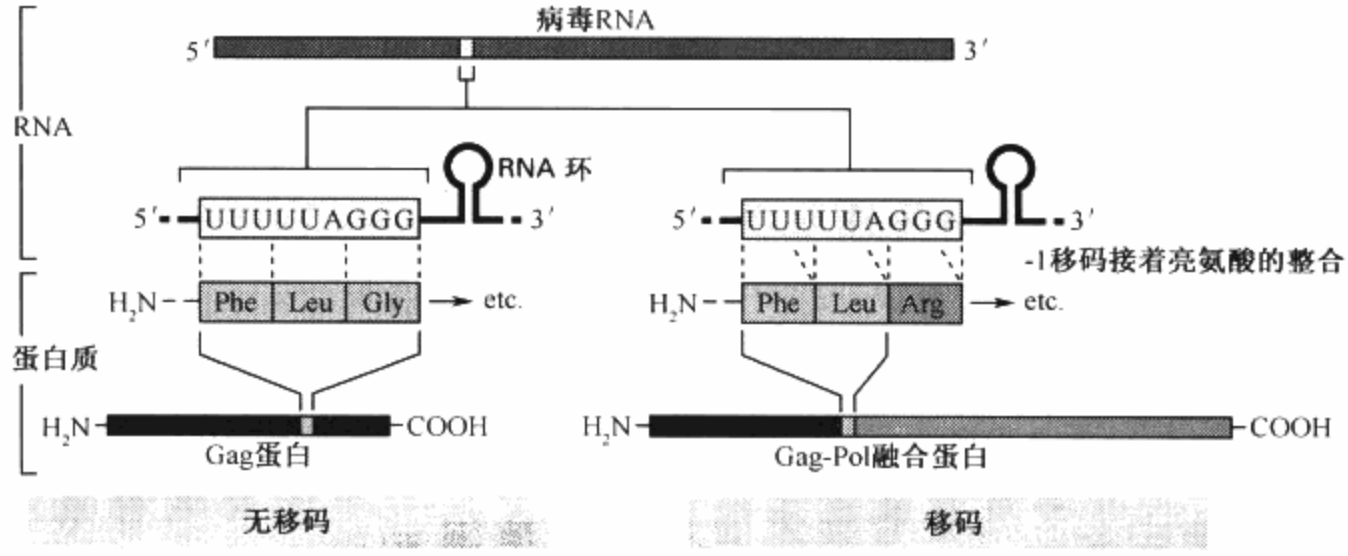


图 6-78　一种反转录病毒合成反转录酶和整合酶的过程中的翻译移码

经过一个由 Gag 和 Pol 氨基酸序列组成的大蛋白质（Gag-Pol 融合蛋白）的酶解加工，合成出病毒反转录酶和整合酶。病毒衣壳蛋白由更丰富的 Gag 蛋白经蛋白酶解加工而来。Gag 和 Gag-Pol 融合蛋白同时开始合成，但 Gag 蛋白在一个框架内终止密码子停止合成（未显示）；图中所示的移码避开了这个终止密码子，从而能够合成更长的 Gag-Pol 融合蛋白。发生移码是因为局部 RNA 结构特征（包括图中所示的 RNA 环）引起附着到生长多肽链末端的 $tRNA^{Leu}$ 偶尔在核糖体向后滑动一个核苷酸，这样它就与 UUU 密码子配对，而不是与最初指令它整合的 UUA 密码子配对；在新可读框中下一个密码子（AGG）指令精氨酸，而不是甘氨酸。这种受调控的滑动一部分原因是因为病毒 mRNA 中形成的茎环结构，如图所示。图中所示的序列来自人类艾滋病病毒，HIV

（改编自 T. Jacks et al., *Nature* 331：280～283，1988）

编码，通过此过程，能够从单个 mRNA 合成不止一种蛋白质。这些病毒一般从同一转录物合成衣壳蛋白（Gag 蛋白），病毒反转录酶以及整合酶（Pol 蛋白）（图 5-73）。这种病毒对 Gag 蛋白的需要量比 Pol 蛋白大得多，并通过在 *gag* 基因后面紧挨着编码 *pol* 基因，但是位于不同的可读框，来实现数量上的调整。在 *gag* 编码序列末端的一个终止密码子上游主动发生的翻译移码能够偶尔避开这个终止密码子。这种移码发生在 mRNA 中的一个特殊密码子上并需要一种特定的再编码信号，它似乎是该位点下游 RNA 序列的一种结构特征（图 6-78）。

许多原核生物蛋白质合成的抑制剂是有用的抗生素

现代医药所有的最有效的抗生素中，许多都是真菌产生的复合物，它们通过作用于细菌蛋白质合成而发挥效果。这些药物中有许多利用了细菌与真核生物核糖体的结构和功能差异，以便优先干扰细菌核糖体的功能。因此这些复合物中有些可以大剂量服用，却对人类没有过度的毒性。由于不同的抗生素结合细菌核糖体不同区域，所以它们常常抑制蛋白质合成过程中不同的步骤。表 6-3 中列出了一些比较常见的这类抗生素以及蛋白质合成的几种其他抑制剂，其中有些作用于真核细胞，因而不能用作抗生素。

表 6-3　蛋白质或 RNA 合成的抑制剂

抑制剂	特定的作用
仅于细菌作用	
四环霉素	阻断氨酰基-tRNA 结合至核糖体的 A 位点
链霉素	阻止由起始复合体到链延长核糖体的转变，同时亦造成编码错误
氯霉素	阻止肽基转移酶在核糖体的反应(图 6-65 的第二步)
赤霉素	阻止核糖体的易位反应(图 6-65 第三步)
利福霉素	结合至 RNA 聚合酶上阻止 RNA 链起始(阻止 RNA 合成)
于细菌及真核生物作用	
嘌呤霉素	加入成长链的末端造成新生肽链未成熟释放
放射菌素力	结合在 DNA 上阻止 RNA 聚合酶移动(阻止 RNA 合成)
作用在真核生物、而不作用在细菌上	
环己酰亚胺放线菌酮	阻止核糖体的易位反应(图 6-65、第三步)
茴香霉素	阻止肽基转移酶在核糖体的反应(图 6-65、第二步)
α-鹅膏菌素	优先结合至 RNA 聚合酶Ⅱ、而阻止 mRNA 的合成

注：真核细胞线粒体(及叶绿体)和细菌在对抑制物的敏感性上相似，因此有些抗生素对人类线粒体有毒害。

由于它们阻断从 DNA 到蛋白质过程中的特定步骤，所以表 6-3 中所列的许多复合物对细胞生物学研究很有用。在这类实验研究中最常用的药物有氯霉素、环己亚胺和嘌呤霉素，它们都特异性地抑制蛋白质合成。例如，在真核细胞中，氯霉素只抑制线粒体内（以及植物的叶绿体内）核糖体上的蛋白质合成，这可能反映出这些细胞器起源于原核生物（讨论见第 14 章），相比之下，环己亚胺只影响细胞质内的核糖体。嘌呤霉素特别有意思，因为它是连接着氨基酸的 tRNA 的结构类似物，因而是另一个分子模拟的实例；核糖体误将其当成真正的氨基酸，并把它整合到生长多肽链 C 端，从而引起多肽链过早成熟和释放。可以想到，嘌呤霉素抑制原核和真核生物中的蛋白质合成。

描述了翻译过程本身之后，现在我们讨论它的产物（细胞的蛋白质）怎样折叠成正确的三维构象。

蛋白质还在合成时就已开始折叠

当利用遗传密码合成出了组成蛋白质的氨基酸序列以后，基因表达的过程还未结束。为了能为细胞所用，这些新合成的多肽链必须折叠起来，形成它独特的三维构象，结合上任何活性必需的小分子辅助因子，再由蛋白质激酶或者蛋白质修饰酶进行适当的修饰，并与同它一起作用的其他蛋白质亚基正确组装起来（图 6-79）。

上面所列的所有蛋白质成熟步骤需要的信息最终包含在相连的氨基酸序列中，它们由核糖体在将 mRNA 分子翻译成多肽链时产生。第 3 章讨论过，当一种蛋白质折叠成紧密的结构

时，它将大部分疏水基团埋入内部核心中。此外，还在分子不同部分之间形成大量非共价相互作用，正是所有这些能量上有利的组织形式的总和决定了多肽链最后的折叠模式——作为最低自由能构象。

当多肽链 N 端开始不断伸出核糖体时，无论是为了它采用的构象，还是为了能迅速地折叠，在数百万年的进化历程中，每个蛋白质的氨基酸序列都经历了选择。许多实验已证明，一旦多结构域蛋白质中有一个结构域从核糖体合成出来，它就在几秒钟内形成一个密实的结构，含有大多数以大致正确的方式排列的最终二级结构（α 螺旋和 β 折叠）（图 6-80）。对于许多蛋白质结构域来说，将这种通常是开放而且可变的结构称作熔球体，它是一个相对较慢过程的起点，在此过程中，许多侧链经过调整最终形成正确的三级结构。然而，由于合成一个大小的蛋白质需要几分钟，所以许多折叠过程在核糖体释放该蛋白质 C 端时就已经完成了（图 6-81）。

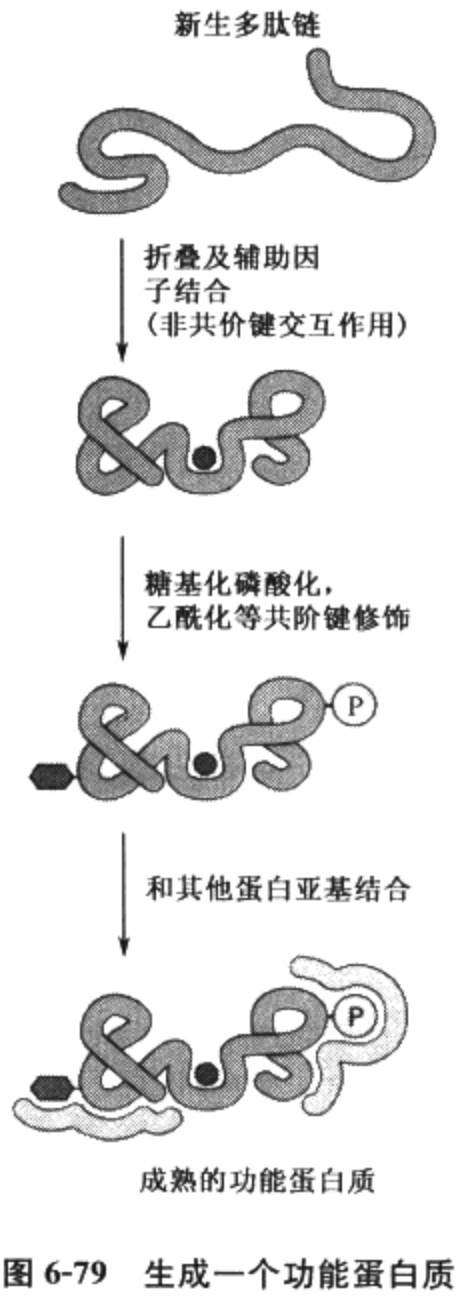


图 6-79　生成一个功能蛋白质经历的步骤

如图所示，mRNA 序列在核糖体上翻译成氨基酸序列并不是形成蛋白质过程的终点。为了对细胞有用，完整的多肽链必须正确地折叠成它的三维构象，结合上任何必需的辅助因子，并与辅助蛋白质一起组装（如果有的话）。这些变化受非共价键形成的推动。如图所示，许多蛋白质也在经过选择的氨基酸上发生共价修饰。虽然最频繁的修饰是蛋白质糖苷化和蛋白质磷酸化，但已知有超过 100 种不同的共价修饰类型（图 4-35）

分子伴侣帮助指导许多蛋白质的折叠

由于一类特殊蛋白质（**分子伴侣**）的存在，所以许多蛋白质折叠就更加有效了。分子伴侣之所以对细胞有用，是因为存在各种不同的途径将一种蛋白质的熔球体形式转变成蛋白质最终紧密的构象。对于许多蛋白质而言，如果没有重新设定折叠过程的分子伴侣的干涉，在此过程中形成的一些中间产物将聚集起来，并以一种偏离正途的形式而存留下来（图 6-82）。

对不允许 λ 噬菌体在其中进行复制的大肠杆菌突变株进行研究后最先鉴定出了分子伴侣。这些突变株产生的分子伴侣装置稍有改变，结果在病毒蛋白质组装的特定步骤存在缺陷。分子伴侣属于热激蛋白（因此将它们命名为 HSP），因为细胞短暂接触升高的温度后它们的量就会显著地增加（如通常在 37℃ 生活的细胞放到 42℃ 中），这反映出某种反馈系统在运行，通过增加分子伴侣的合成量，帮助错误折叠的蛋白质（如升高温度后合成的蛋白质）重新进行折叠，从而对这些蛋白质的任何一点增加做出反应。

真核细胞至少有两个主要的分子伴侣家族：hsp60 和 hsp70 蛋白。不同的家族成员在不同的细胞器发挥作用。因此，正如第 12 章讨论的，线粒体含有它们自己的 hsp60 和 hsp70 分子，它们与细胞质中发挥作用的那些伴侣蛋白不同，内质网中还存在一种特殊的 hsp70（称作 BIP）帮助折叠蛋白质。

在帮助其他蛋白质折叠时，hsp60 类似蛋白质和 hsp70 蛋白各自与自己的少量结合蛋白质共同发挥作用。它们对未完全折叠蛋白质裸露的疏水结构区域有共同的亲和力，并水解 ATP，每次 ATP 水解循环通常都结合或释放了它们的蛋白质。在其他方面，两种 hsp 蛋白功能也不同。hsp70 机制在蛋白质生命早期就开始发挥作用，蛋白质离开核糖体之前，就已

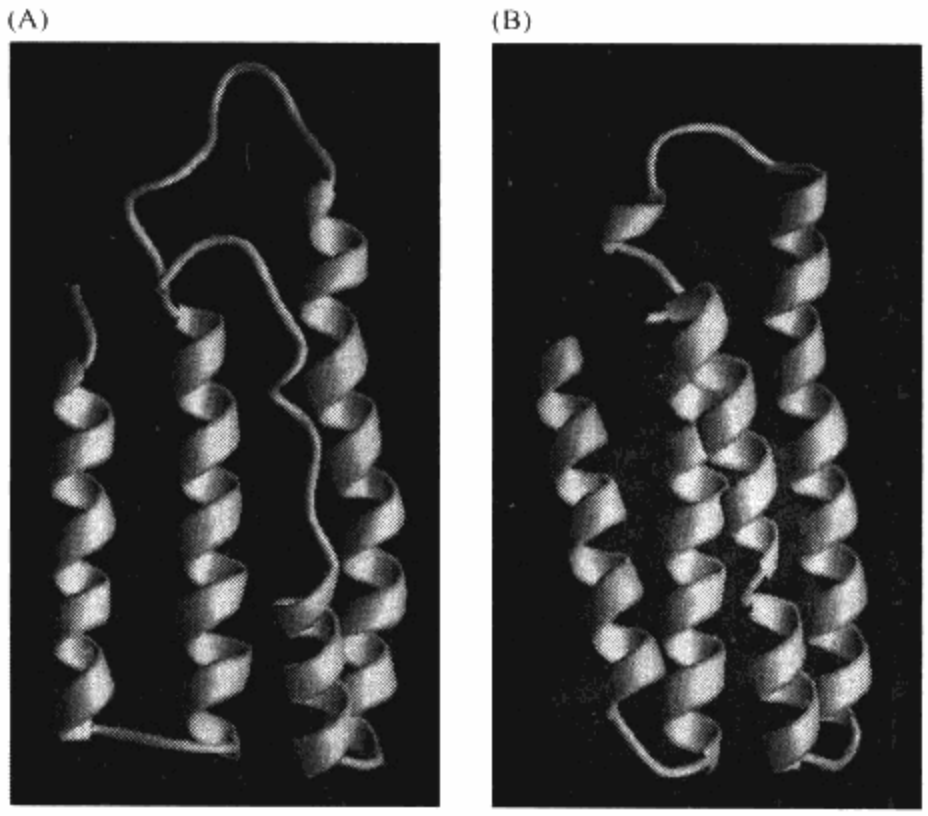


图 6-80　熔球的结构

（A）细胞色素 562b 的熔球形式比（B）中显示的该蛋白质最终折叠形式更加开放，也不那么高度有序。注意熔球含有大多数二级结构的最终形式，尽管 α 螺旋的末端散开了并且其中一个螺旋只形成了一部分(经 Joshua Wand 允许后使用，摘自 Y. Feng et al.,*Nat. struct. Biol.* 1：30～35，1994)

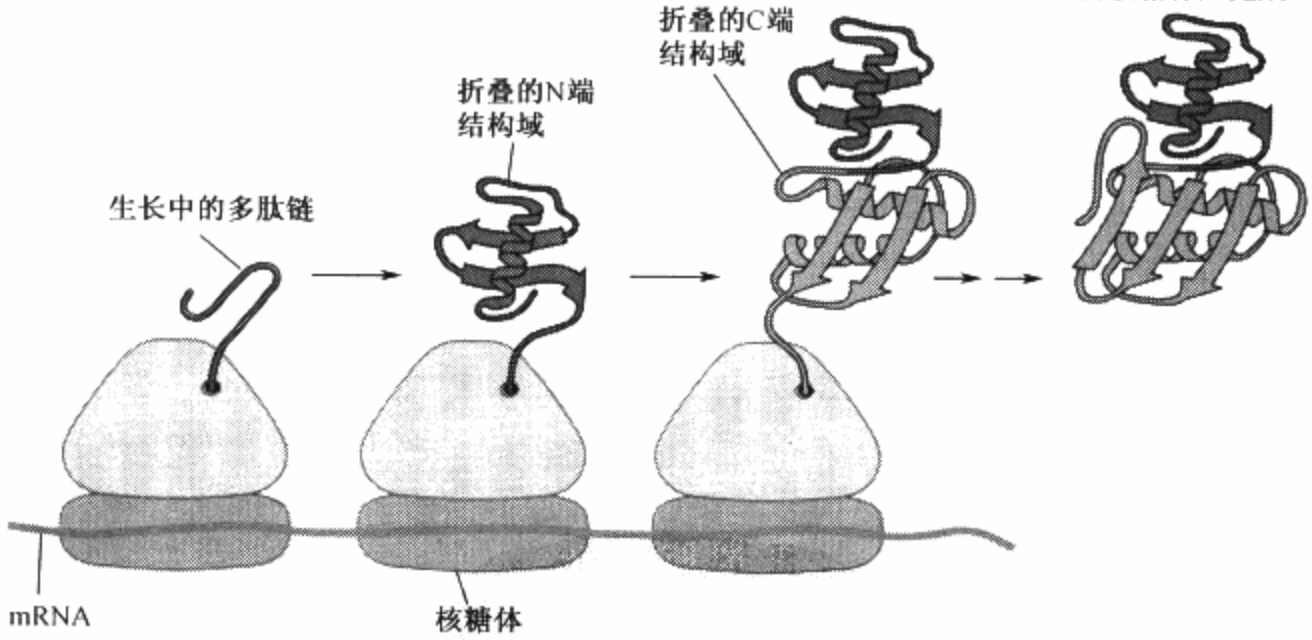


图 6-81　蛋白质的共翻译折叠

图中所示的生长多肽链刚从核糖体上出现时就获得其二级和三级结构。N 端结构域首先折叠起来，此时 C 端仍然在合成。在这种情况下，到此时为止还没有获得最终构象的蛋白质从核糖体上释放下来（摘自 A. N. Federov and T. D. Baldwin. *J. biol. chen.* 272：32715～32718，1997）

结合了一串大约 7 个疏水的氨基酸（图 6-83）。相比之下，hsp60 类似蛋白质形成一个巨大的桶状结构，在蛋白质生命晚期，即它已完全合成出来之后发挥作用。这种伴侣形成一种“隔离室”，将错误折叠的蛋白质装起来，防止它们积聚，并提供一种有利环境让它们尝试重新折叠（图 6-84）。

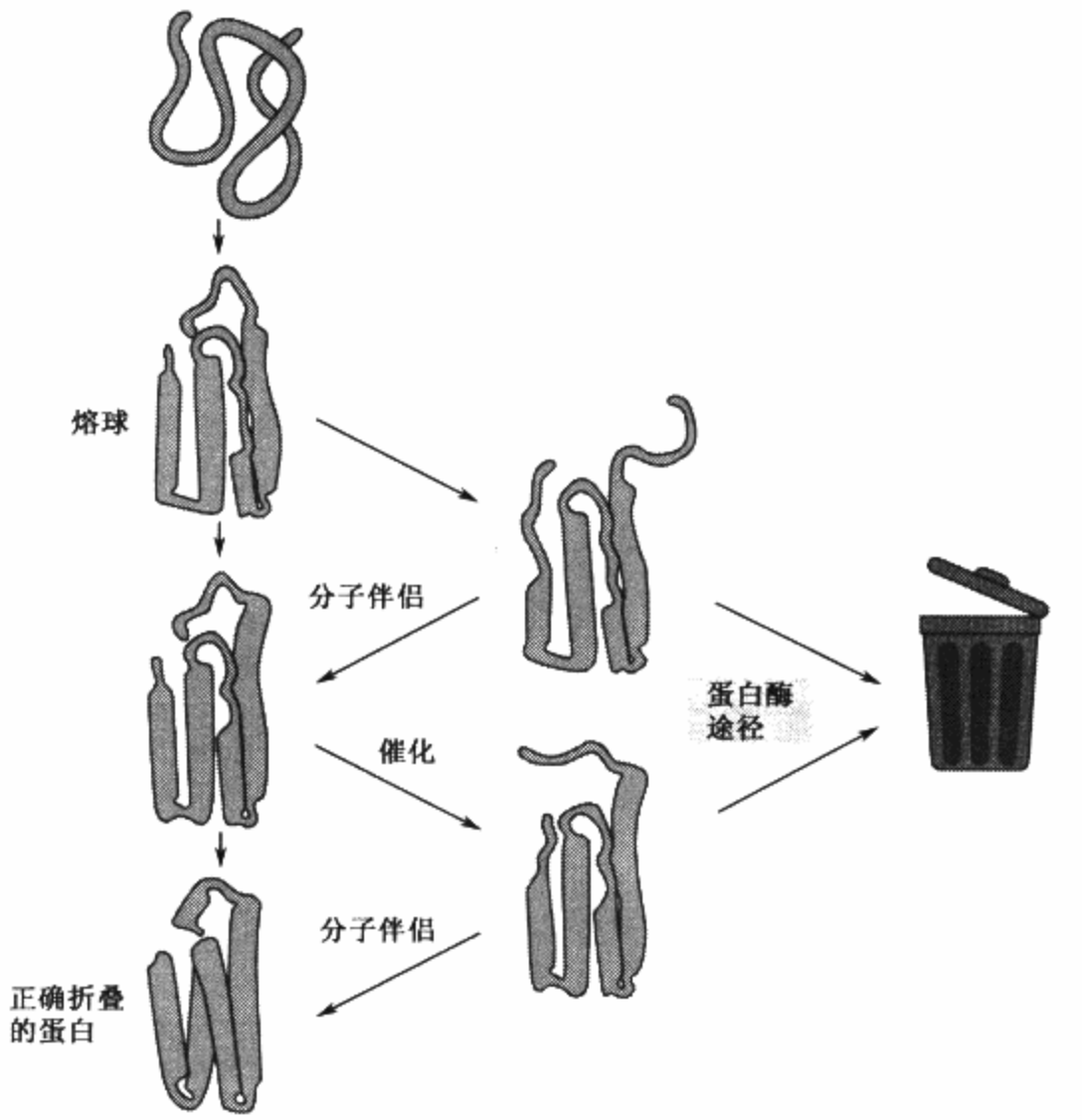


图 6-82　目前有关蛋白质折叠的观点

新合成蛋白质的每个结构域迅速到达“熔球”状态。随后的折叠更加缓慢且经由多条途径，常常涉及分子伴侣的帮助。有些分子可能仍然不能正确折叠；它们由特殊的蛋白酶识别和降解，这在前面解释过

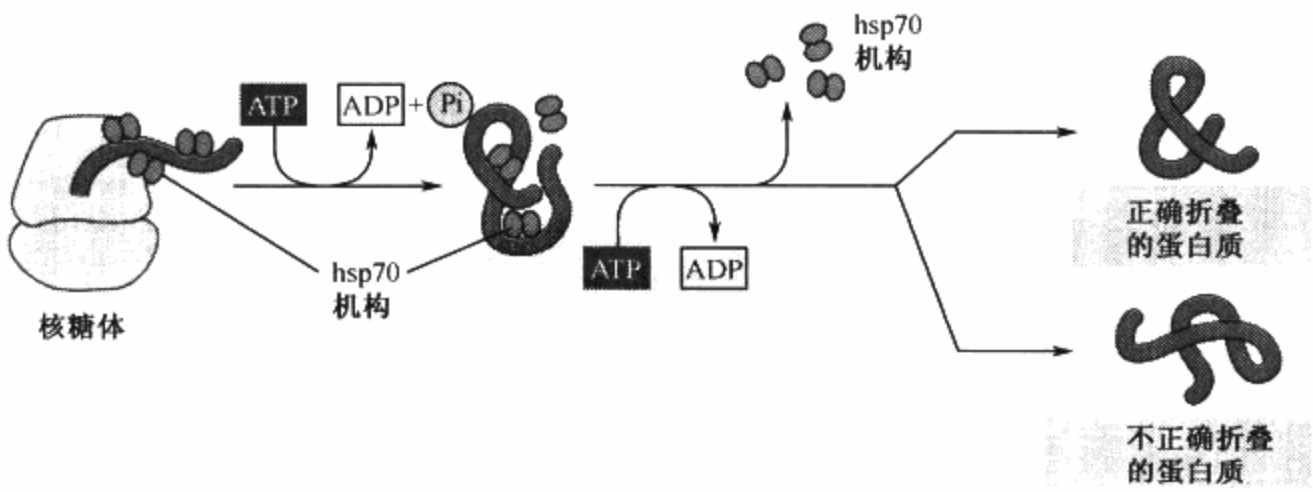


图 6-83　hsp70 分子伴侣家族

这些蛋白质在蛋白质折叠早期发挥作用，识别蛋白质表面一小段疏水氨基酸。在一组更小的 hsp40 蛋白帮助下，一个 hsp70 单体结合它的目标蛋白质，然后将一分子 ATP 水解成 ADP，从而发生构象变化，使 hsp70 非常牢固地钳住目标蛋白质。hsp40 解离以后，ATP 在释放 ADP 后迅速地重新结合，诱导 hsp70 蛋白解离。hsp 蛋白迅速的结合和释放循环帮助目标蛋白质再折叠，如图 6-84 中的概括说明

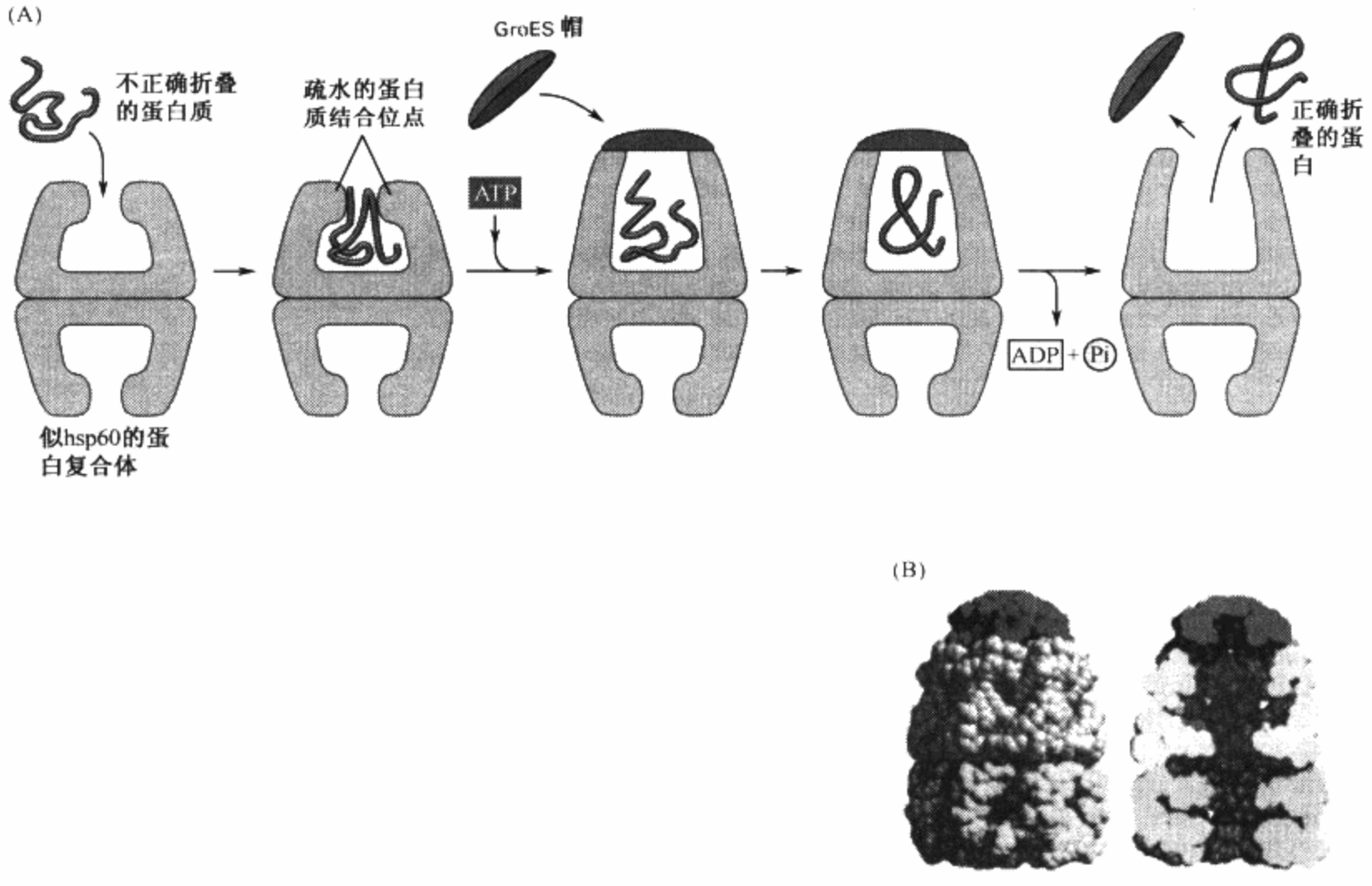


图 6-84　hsp60 分子伴侣家族的结构与功能

（A）蛋白质再折叠的催化。如图所示，错误折叠的蛋白质首先由一个桶边缘上的疏水相互作用捕获，随后 ATP 以及一个蛋白质帽子的结合增加了桶边缘的直径，这可以短暂地打开（部分解折叠）进入的蛋白质。同时这也将蛋白质限制在封闭的空间内，在这里获得新的折叠机会。大约 15s 以后，ATP 水解作用排出蛋白质，无论它是否折叠，然后重复循环。这类分子伴侣也称作伴侣蛋白，在线粒体中命名为 hsp60，在脊椎动物细胞质中叫做 TCP-1，在细菌中叫做 GroEL。如图所示，任一时刻该对称桶只有一半作用于进入的蛋白质。（B）结合到 GroES 帽子上的 GroEL 结构，利用 X 射线晶体学得到。左边显示的是桶状结构的外侧，右边是其中央纵切面图[(B)摘自 B. Bukace, and A. L. Horwich, *Cell* 92：351～366，1998]

暴露的疏水区域为蛋白质质量控制提供了关键信号

如果短时间内将放射性氨基酸加入到细胞内，就可以追踪新合成的蛋白质直至它们最终功能形式。就是这种实验表明了 hsp70 蛋白最先发挥作用，当蛋白质还在核糖体上进行合成时，就已经开始了，而 hsp60 类似蛋白质只在晚期用于帮助折叠完整的蛋白质。可是相同的实验也揭示出新合成的蛋白质只用一部分参与了这个过程；可能所有蛋白质大约有 20%具有 hsp70 分子伴侣，10%有 hsp60 类似分子伴侣。怎样对这些蛋白质进行筛选后用于这种 ATP 催化的重折叠过程呢？

在回答这个问题之前，我们需要暂停一下，来更广泛的考虑蛋白质的翻译后命运，在表面暴露着相当大的疏水氨基酸区域的蛋白质通常是异常的：它或者在离开核糖体之后不能正确折叠，随后会发生部分解折叠，或者不能找到更大的蛋白复合物中正常的伙伴亚基。这种蛋白质不仅仅是对细胞无效，还可能有害。许多具有异常暴露疏水区域的蛋白质能发生聚积，并在溶液中析出。我们应当看到，在少见的情况下，确实发生了这种聚积并引起严重的人类疾病。但在大多数细胞内，强大的蛋白质量控制机制防止了这种灾难的发生。

考虑到这种背景，就毫不奇怪细胞会进化出复杂的机制，识别并消除蛋白质上的疏水区域了。这些机制中有两种依赖于刚才讨论的分子伴侣，它们结合到该区域，通过给缺陷蛋白质再次折叠的机会尝试对它进行修复。同时通过覆盖疏水区域，这些伴侣蛋白瞬间防止了蛋白质聚积。自身能迅速而又正确地折叠的蛋白质并不会展示出这些区域，因而被伴侣蛋白略过。

图 6-85 概括了细胞为一个难以折叠，新合成的蛋白质做出的所有质量控制选择。如图所示，如果尝试失败了，就会引起第三种机制发挥作用，由蛋白质水解作用将该蛋白质彻底降解掉。水解途径起始于蛋白质表面异常疏水区域的识别，结束于将整个蛋白质呈递到一种蛋白质破坏装置，它是一种称作蛋白质酶体的复杂蛋白酶。下面将会描述到，这个过程依赖于一个复杂的蛋白标记系统，该系统也通过在细胞内破坏选择出来的正常蛋白质行使其他重要功能。

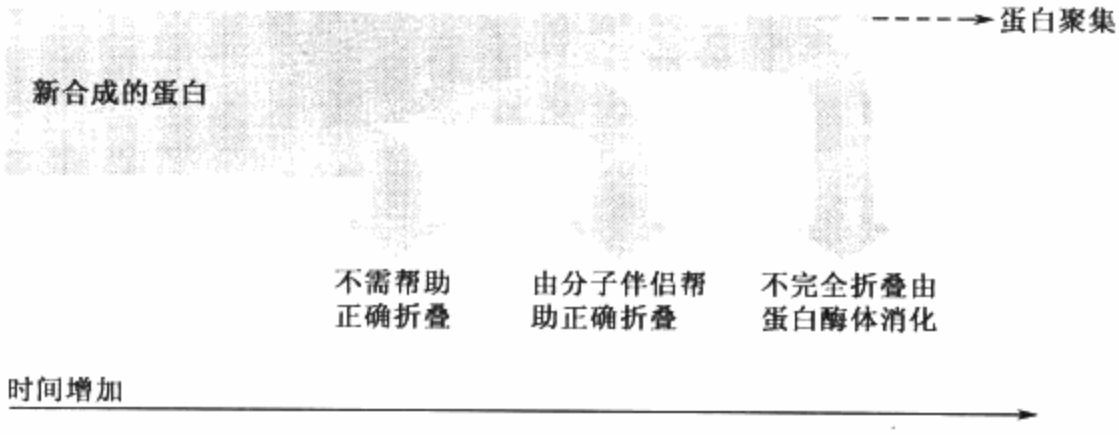


图 6-85　蛋白质合成后监控蛋白质量的细胞机制

如图所示，新合成的蛋白质有时正确地折叠起来并自由决定与其辅助蛋白组装，在这种情况下它单独存在。不完全折叠的蛋白质由分子伴侣帮助重新折叠：首先由 hsp70 蛋白提供帮助，如果失败，接着由 hsp60 类似蛋白质提供帮助。在两种情况下，进入的蛋白质由它们表面异常裸露出的疏水区识别。这些过程与一个不同的系统竞争，该系统识别异常裸露区并将含有它的蛋白质转移到蛋白酶体中进行彻底的降解。细胞中防止大量蛋白质聚集需要所有这些过程组合起来发挥作用，蛋白质上面许多疏水区凝集在一起时会发生这种聚集并从溶液中成块沉淀出来

蛋白酶体降解细胞内大部分新合成的蛋白质

细胞迅速地清除了翻译过程的失败产物。最新的实验研究表明，作为刚才描述的蛋白质质量控制的结果，多达 1/3 新合成的多肽链被选择出来迅速降解掉。真核生物中最终清除装置是**蛋白酶体**，它是一种广泛存在的依赖 ATP 的蛋白酶，大约占细胞蛋白质的 1%。该蛋白酶体以大量的拷贝数分散存在于细胞质和细胞核内，还作用于内质网的蛋白质（ER）：那些进入 ER 后或者不能折叠，或者不能正确组装的蛋白质，由基于 ER 的监视系统识别，并将它们反转移到细胞质中进行降解（讨论见第 12 章）。

每个蛋白酶体呈一个中空的圆柱体（20S 核心蛋白酶体），含有多个蛋白质亚基构成，它们以 4 个七聚体环堆叠成圆柱体而组装起来。这些亚基中有些是不同的蛋白酶，其活性位点面向圆筒内腔[图6-86(A)]。该圆筒两端通常结合有一个巨大的蛋白质复合物（19S 帽子），含有将近 20 种不同的多肽[图6-86(B)]。帽子亚基包括至少 6 种水解 ATP 的蛋白质；这些 ATP 酶位于圆筒边缘附近，通常认为它们能将蛋白酶折叠，以便用于消化，并将它们转移至内腔用于蛋白质水解。蛋白酶体的关键功能，也是它自身结构复杂的原因之一，就是其机制

的持续作用的能力：与“简单”的蛋白酶在解离之前只切割多肽一次相比，蛋白酶体使整个底物保持结合状态，直到它全部转变成短肽片段为止。

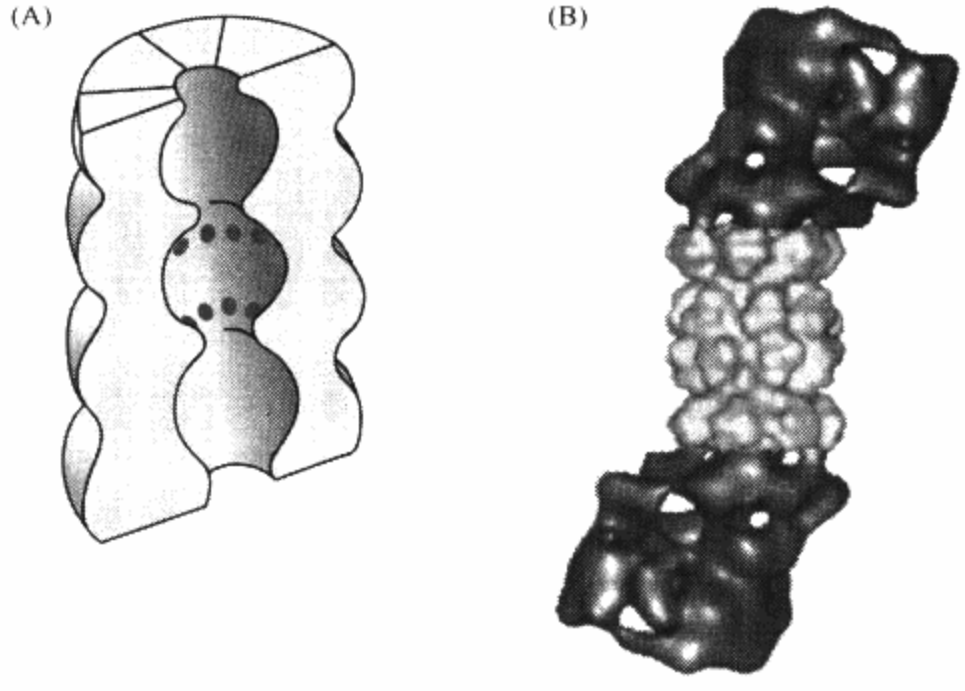


图 6-86　蛋白酶体

（A）利用 X 射线晶体学得到的 20S 中央圆筒结构剖面图，红点表示蛋白酶活性位点。（B）整个蛋白酶体的结构，其中中央圆筒（黄色）两个末端各添加一个 19S 帽子，其结构已用电镜成像电脑处理得到。这个复杂帽子选择性地结合那些已标记过的要降解的蛋白质；然后它利用 ATP 水解使它们的肽链解折叠，并将它们放到 20S 圆筒内部空间降解成小肽段[(B)摘自 W. Baumeister et al., *Cell* 92：367～380，1998. © Elsevier]

19S 帽子作为进入内部酶解腔室入口处的调节“大门”，也负责将目标蛋白质底物结合到蛋白酶体上。除少数情况例外，蛋白酶体作用于那些已经由多个称作泛素的共价相连小蛋白质特异性标记用于降解的蛋白质[图6-87(A)]。泛素在细胞内或者游离存在，或者共价结合到各种细胞内蛋白质上。对于大多数这种蛋白质而言，由泛素进行的这种标记导致它们被蛋白酶体降解掉。

一种复杂的泛素结合系统标记将要降解的蛋白质

泛素用于和其他蛋白质在 ATP 依赖的泛素激活酶（E1）的催化下进行结合，这种酶作用后生成一个活化的泛素蛋白，它再转移到一组泛素结合酶（E2）其中一个上面。E2 酶与辅助蛋白质（E3）联合作用。在称为泛素连接酶的 E2-E3 复合体中，E3 组分结合到蛋白质中特殊的降解标记上，从而帮助 E2 形成与底物蛋白质赖氨酸相连的多泛素链。在该链中，每个泛素 C 端残基与前面泛素上面特定的赖氨酸相连，形成了一个线性的泛素-泛素连接[图6-87(B)]。正是靶蛋白质上这种多泛素链，由蛋白酶体中特异的受体进行识别。

哺乳动物中存在大约 30 种结构相似但不同的 E2 酶，以及和特定的 E2 酶形成复合体的数百种不同的 E3 蛋白。因此泛素蛋白酶体系统由许多不同、但组织形式相似的蛋白质水解途径组成，该途径普遍在“上部”有一个 E1 酶，“底部”有一个蛋白酶体，并在 E2-E3 泛素连接酶和辅助因子的组成方面不同。不同的泛素连接酶识别不同的降解信号，并因而对细胞内具有这些信号的不同蛋白质部分进行降解。

变性的或其他情况下错误折叠的蛋白质，以及含有已氧化或其他异常氨基酸的蛋白质，被识别出来并破坏掉，因为异常蛋白质倾向于在表面呈现出氨基酸序列或构象模式，它们作为降解信号，由泛素-蛋白酶体系统中一组 E3 分子进行识别；在正常的这类蛋白质中，这些

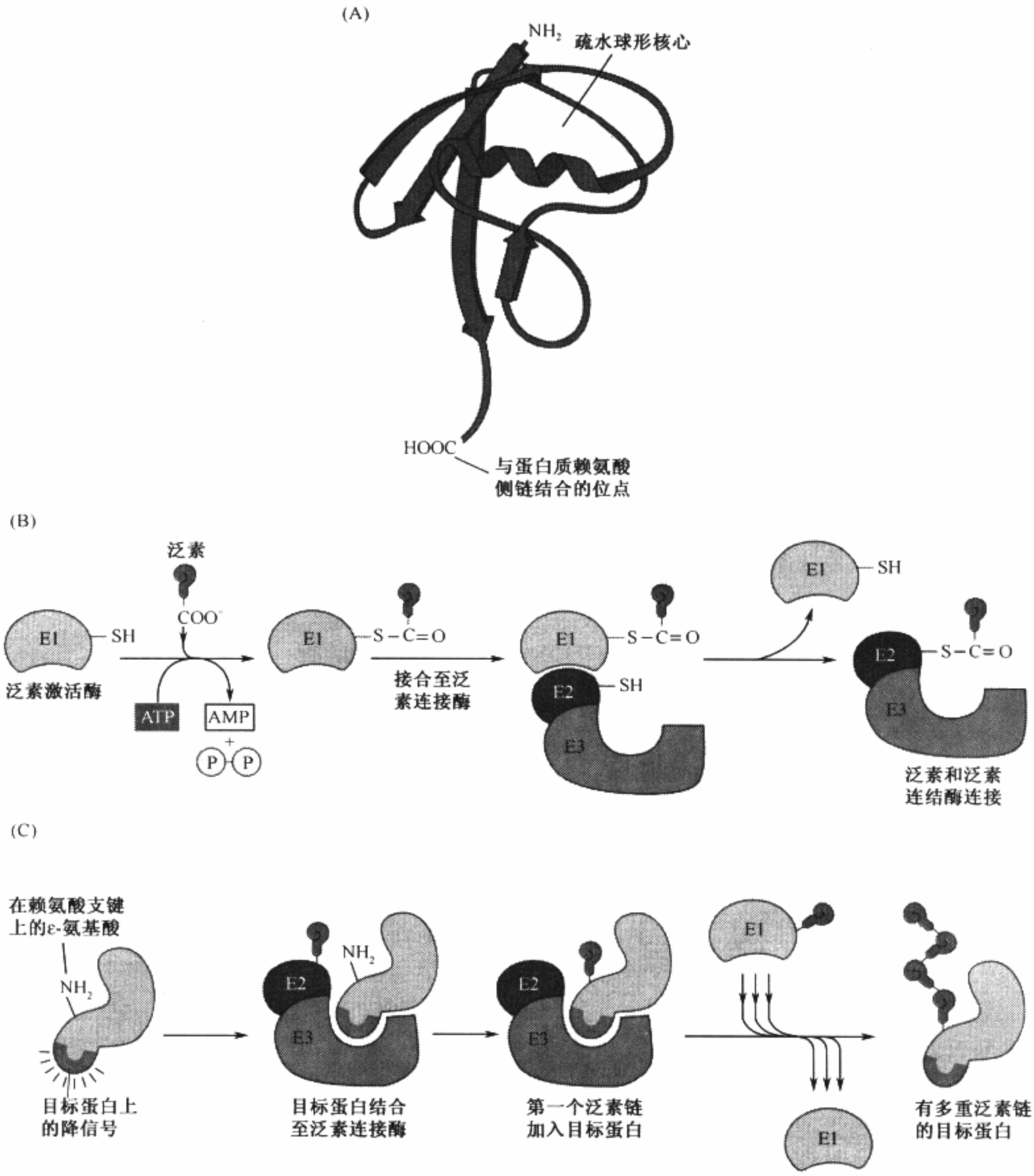


图 6-87　泛素以及用多泛素链标记蛋白质

(A) 泛素的三维结构，这个相对较小的蛋白质含有 76 个氨基酸。(B) 泛素的 C 端最初与 E1 蛋白上的半胱氨酸侧链形成一个高能硫酯键而活化。该反应需要 ATP，并经过一个共价 AMP-泛素中间产物。E1，也称作泛素活化酶，上面活化的泛素然后转移到一组 E2 分子的半胱氨酸上。这些 E2 与更大的 E3 分子家族形成复合物。(C) 目标蛋白质上增加一条多泛素链。哺乳动物细胞中大约有 300 种不同的 E2-E3 复合物，它们每个都通过 E3 组分识别目标蛋白质上不同的降解信号。E2 称作泛素结合酶。传统上将 E3 称为泛素连接酶，但是这样称呼有功能的 E2-E3 复合物会更加精确一些

序列当然必须深埋起来，因而不能接触到。然而，识别并破坏异常蛋白质的蛋白质水解途径，必须能够区分具有错误构象的完整的蛋白质，及还未实现正常折叠构象的仍在核糖体上生长的许多多肽（以及刚从核糖体上释放的多肽）。这并不是一个无关紧要的问题；泛素-蛋白酶

系统据认为会破坏某些新生的和刚合成的蛋白质分子，不是因为这些蛋白质像这样异常，而是因为它们短暂地暴露了深埋在成熟（折叠）状态中的降解信号。

许多蛋白质受到可调型破坏的控制

刚才描述过，细胞内蛋白质水解机制的功能之一就是识别和清除错误折叠以及不正常的蛋白质。但是这些蛋白质水解途径另一个功能却是赋予特定的正常蛋白质以短半衰期，它们的浓度必须随细胞所处状态的改变而迅速变化。这些短寿命蛋白质有些在任何时候都迅速被降解，而其他许多只是有条件地短寿命，即在某种条件下，它们在代谢上是稳定的，如果细胞状态发生改变，就变得不稳定了。例如，有丝分裂周期蛋白在整个细胞周期寿命都是长寿命的，一旦到达有丝分裂结束时就突然降解掉，这些将在第 17 章解释。

怎样控制蛋白质的这种受调节的破坏呢？已知存在多种不同的机制，本书后面会就这一方面通过特定例子给予说明。在一类通用机制中[图6-88(A)]，E3 的磷酸化或者 E3 蛋白结合一种特殊的小或大分子引起它的别构转变，打开了泛素连接酶的活性。例如，促后期复合体（APC）是一种多亚基泛素连接酶，在有丝分裂中，由于与细胞周期同步的亚基的加入而激活。然后激活的 APC 引起有丝分裂周期蛋白和其他多种中期-后期转变调节因子的降解(图 17-20)。

作为选择，在应答细胞内信号或者环境中信号时，蛋白质内可以产生一种降解信号，引起自身迅速泛素化，并由蛋白酶体破坏掉。一种常见的产生这种信号的方式就是将蛋白质的一个特殊位点磷酸化，从而暴露出正常情况下深埋起来的降解信号。另一种暴露这种信号的方式就是某个蛋白质亚基可调节性地解离。最后，假如肽键的一次切割产生一个新的 N 端，并由特定的 E3 识别成“去稳定的”N 端残基[图6-88(B)]，也可以产生强大的降解信号。

N 端类型降解信号的出现是因为“N 端规则”，这个规则将蛋白质在体内的半衰期与它的 N 端残基关联起来。20 种标准氨基酸中，有 12 种在芽殖酵母 N 端规则中属于去稳定残基(Arg、Lys、His、Phe、Leu、Tyr、Trp、ILe、Asp、Glu、Asn 和 Gln)。这些去稳定 N 端残基由一种特殊的泛素连接酶识别，从酵母到人类中，它都很保守。

我们已经看到，所有最初合成的蛋白质都有一个甲硫氨酸（或者细菌中的甲酰甲硫氨酸）作为它们的 N 端残基，在 N 端规则中，它是一个稳定的残基。一种称作甲硫氨酸氨肽酶的特殊蛋白酶常常将新生蛋白质的第一个甲硫氨酸除去，但是只有当第二个残基在酵母类型 N 端规则中也属于稳定残基时，它们才会这样做。因此，最初不清楚 N 端规则底物在体内是如何形成的。然而最近已经发现，将姐妹染色单体维系在一起的蛋白质复合物，粘连蛋白的一个亚基在中期-后期转变时受到一个位点特异蛋白酶的切割。这种受细胞周期调节的切割使姐妹染色质分离，并导致有丝分裂的完成（图 17-26）。被切割的亚基 C 端片段拥有一个 N 端精氨酸，它在 N 端规则中是一种去稳定残基。缺少 N 端规则途径的突变体细胞显示出大大增加的染色体丢失概率，这可能是因为无法降解掉这个粘连蛋白亚基片段，因而干扰了下一细胞周期中新的染色质结合粘连蛋白复合体的形成。

异常折叠的蛋白质能够积聚起来，引起破坏性的人类疾病

如果细胞所有的蛋白质量控制失效了，就会在受影响的细胞内形成巨大的蛋白质积聚物(图 6-85)，有些这样的积聚物通过吸收重要的大分子结合到它们上面，能够严重地破坏细胞，甚至引起细胞死亡。组织内从死亡细胞中释放的蛋白质积聚物倾向于在细胞周围的胞外基质中积累，在极端情况下它们也能破坏组织。由于大脑由高度组织起来的神经细胞形成，所以

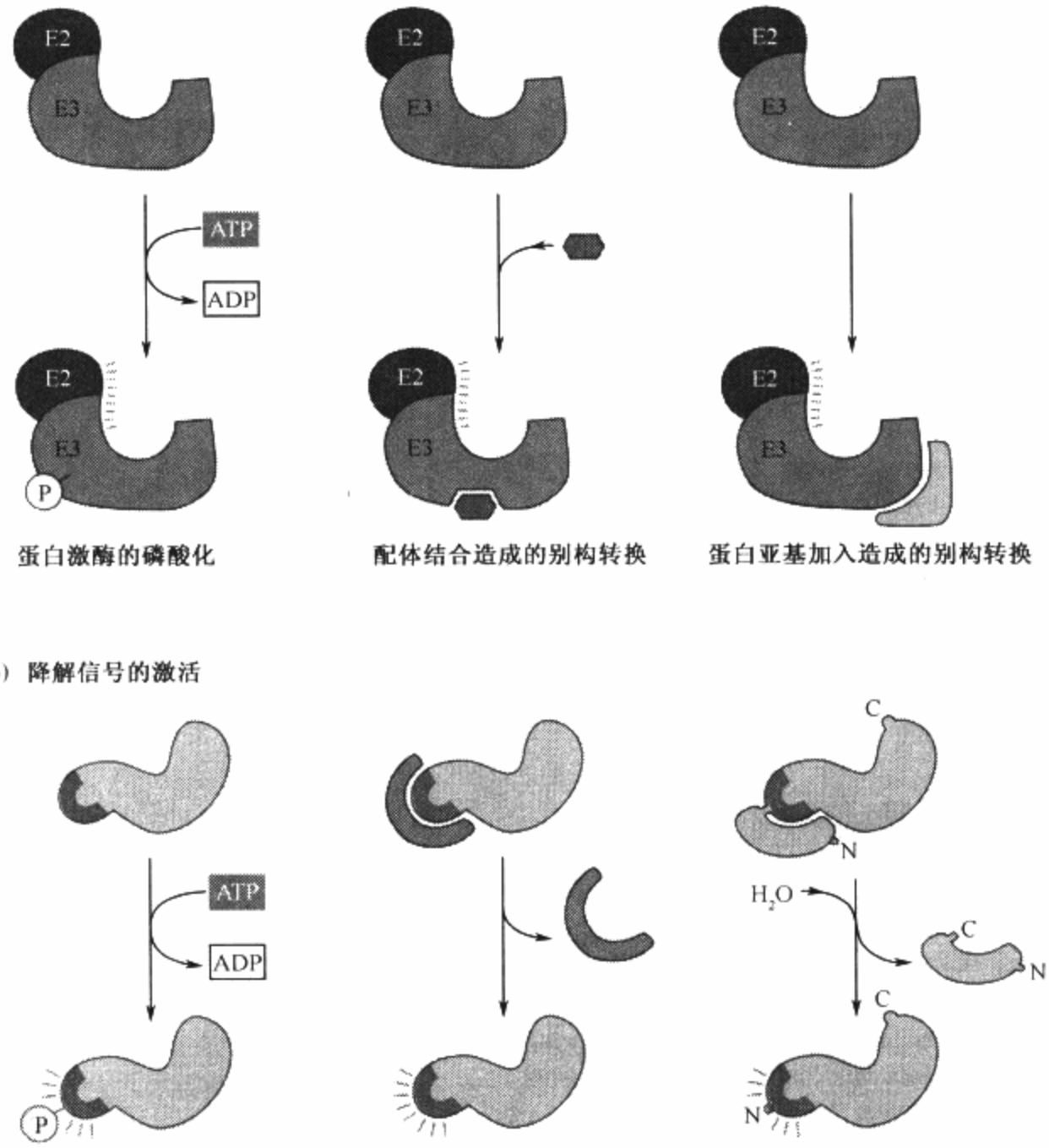


图 6-88　两种诱导特定蛋白质降解的一般方式

(A) 特定 E3 分子的激活产生一个新的泛素连接酶。(B) 将要降解的蛋白质中产生一个裸露的信号，这个信号结合泛素连接酶，从而在目标蛋白质上使附近的一个赖氨酸增加一条多泛素链。已知图中所示的全部 6 条途径都被细胞用于诱导经过选择的蛋白质进入蛋白酶体

它特别容易受伤害。因此，毫不奇怪蛋白质积累主要引起神经变性疾病。其中最著名的就是亨庭顿病和阿尔茨海默病——后者在当今世界 2000 多万人中引起与年龄有关的痴呆。

特定的某种蛋白质积聚要存活，生长并破坏生物体，就必须对细胞外部和内部的蛋白酶解作用具有高度的抵抗力。许多引发问题的蛋白质聚集物能形成原纤维，它们由许多多肽链以连续的 β 折叠一层一层堆叠起来而形成。这种所谓的交叉 β 纤维[图6-89(C)]对蛋白质水解作用倾向于高度抗性，这种抵抗力可能解释了为什么在蛋白质聚集物引起的如此多神经生物学紊乱中都能观测到这种结构，它在这些紊乱情况中产生了异常的着色沉积，称为淀粉样蛋白。

这些疾病中的一种特定类型已经臭名远扬，它们是**朊病毒病**。与亨庭顿病或阿尔茨海默

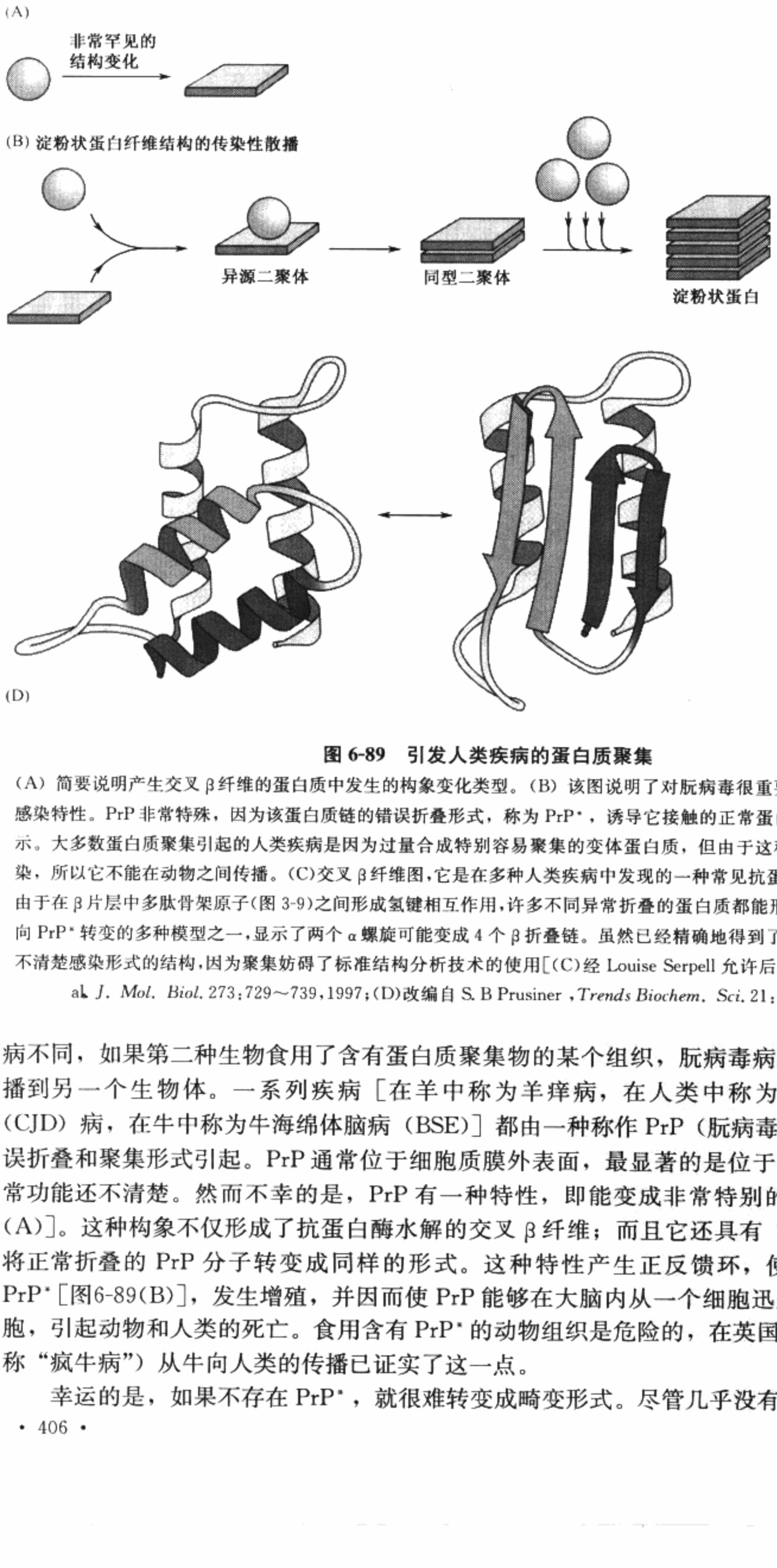


图 6-89　引发人类疾病的蛋白质聚集

(A) 简要说明产生交叉 β 纤维的蛋白质中发生的构象变化类型。(B) 该图说明了对朊病毒很重要的蛋白质聚集的自我感染特性。PrP 非常特殊，因为该蛋白质链的错误折叠形式，称为 PrP^*，诱导它接触的正常蛋白质改变构象，如图所示。大多数蛋白质聚集引起的人类疾病是因为过量合成特别容易聚集的变体蛋白质，但由于这种结构不以这种方式感染，所以它不能在动物之间传播。(C)交叉 β 纤维图，它是在多种人类疾病中发现的一种常见抗蛋白酶的蛋白质聚集体。由于在 β 片层中多肽骨架原子(图 3-9)之间形成氢键相互作用，许多不同异常折叠的蛋白质都能形成这种结构。(D)PrP 向 PrP^* 转变的多种模型之一，显示了两个 α 螺旋可能变成 4 个 β 折叠链。虽然已经精确地得到了正常蛋白的结构，但还不清楚感染形式的结构，因为聚集妨碍了标准结构分析技术的使用[(C)经 Louise Serpell 允许后使用，摘自 M. Sunde et al. *J. Mol. Biol.* 273:729～739,1997;(D)改编自 S. B Prusiner, *Trends Biochem. Sci.* 21:482～487,1996]

病不同，如果第二种生物食用了含有蛋白质聚集物的某个组织，朊病毒病能从一个生物体传播到另一个生物体。一系列疾病［在羊中称为羊痒病，在人类中称为 Creutz feldt-jacob (CJD) 病，在牛中称为牛海绵体脑病 (BSE)］都由一种称作 PrP (朊病毒蛋白) 的蛋白质错误折叠和聚集形式引起。PrP 通常位于细胞质膜外表面，最显著的是位于神经元中。它的正常功能还不清楚。然而不幸的是，PrP 有一种特性，即能变成非常特别的畸形构象[图6-89(A)]。这种构象不仅形成了抗蛋白酶水解的交叉 β 纤维；而且它还具有“感染性”，因为它将正常折叠的 PrP 分子转变成同样的形式。这种特性产生正反馈环，使畸形的 PrP 称作 PrP^*[图6-89(B)]，发生增殖，并因而使 PrP 能够在大脑内从一个细胞迅速传播到另一个细胞，引起动物和人类的死亡。食用含有 PrP^* 的动物组织是危险的，在英国观察到 BSE (通常称“疯牛病”) 从牛向人类的传播已证实了这一点。

幸运的是，如果不存在 PrP^*，就很难转变成畸变形式。尽管几乎没有蛋白质可能错误折

叠成具有感染性的构象，但是已经在酵母细胞中发现类似的转变引起神秘的“蛋白质遗传”。

从 DNA 到蛋白质要经过许多步骤

到现在为止我们在本章已经看到，从包含在基因中的信息合成出一个正确折叠的蛋白质需要许多不同的化学反应（图 6-90）。因此，细胞内正确折叠的蛋白质最终水平依赖于这么多步骤中每一步的效率。

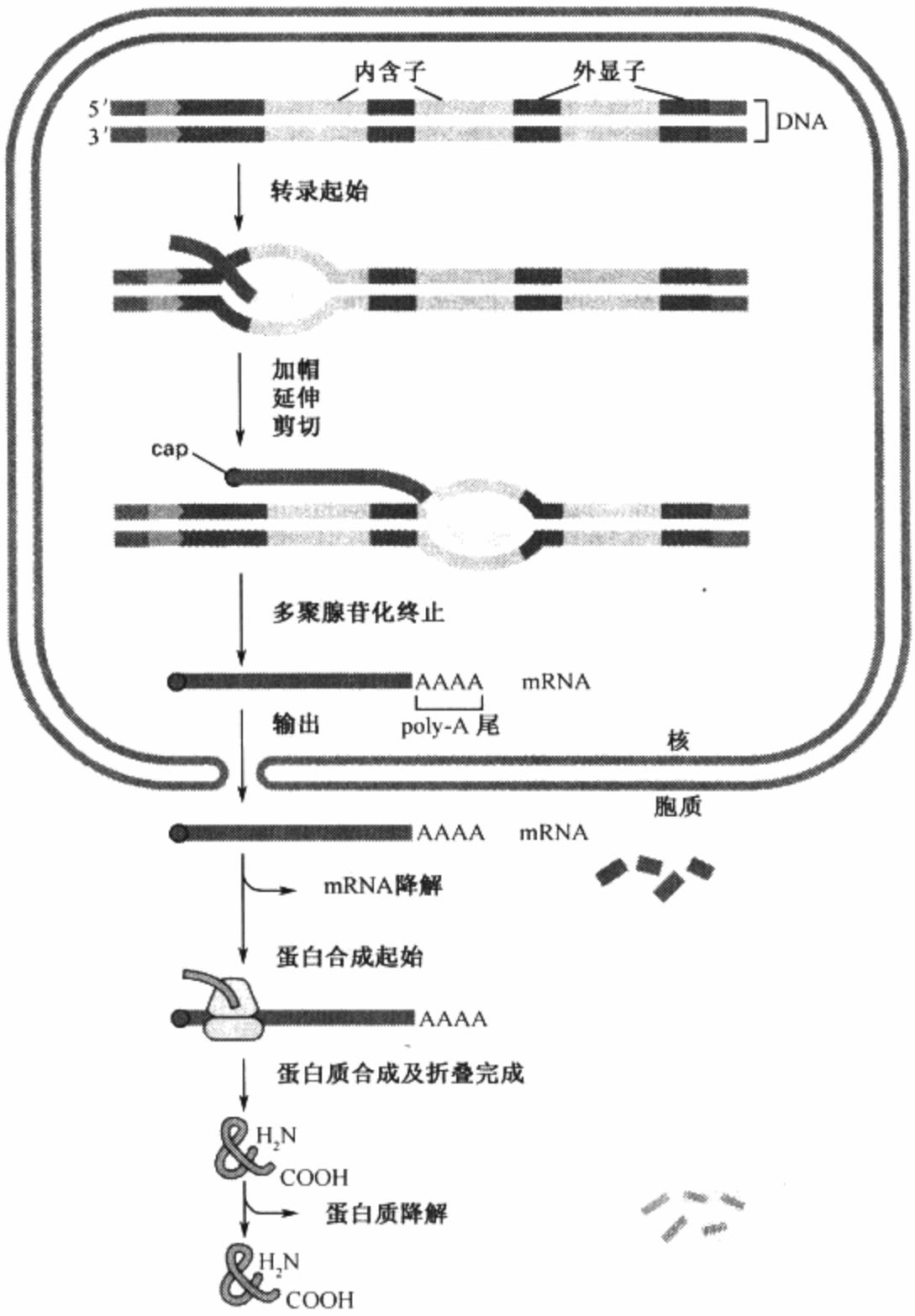


图 6-90　一个真核细胞蛋白质的合成

真核细胞中每个蛋白质最终水平决定于图中所描绘的每个反应步骤的效率

第 7 章我们将讨论细胞拥有根据其需要改变蛋白质水平的能力。从理论上讲，对于每个蛋白质，图 6-90 中的任何或者所有步骤，可以受到细胞的调节。然而，我们在第 7 章会看到，转录起始是细胞调节每个基因表达最常见的点，之所以这样是因为防止基因表达最有效的方式就是阻止第一步（DNA 序列转录成 RNA 分子）的发生。

小结

mRNA 分子的核苷酸序列翻译成蛋白质在细胞质中称作核糖体的巨大核糖核蛋白组装物上进行。用于蛋白质合成的氨基酸首先结合到一类 tRNA 分子上，它们每一个都能通过互补碱基配对相互作用识别 mRNA 中特定的一套核苷酸三联体（密码子）。然后 mRNA 中的核苷酸序列根据遗传密码，以 3 个一组从一端向另一端阅读。

为了起始翻译，核糖体小亚基在起始密码子位置结合上 mRNA 分子，该密码子由一个独特的起始 tRNA 分子识别。核糖体大亚基结合上来以后，形成完整的核糖体，并开始蛋白质合成的延伸阶段。在此阶段中，通过密码子和反密码子之间形成互补碱基配对，氨酰-tRNA（每个拥有一个特殊的氨基酸）依次与 mRNA 中的适当密码子结合。通过循环三个连续步骤的方式，每个氨基酸加到生长多肽链的 C 端：首先氨酰-tRNA 结合上，然后形成肽键，再进行核糖转位。mRNA 分子一个密码接一个密码地沿 5′→3′方向前进，穿过核糖体直到到达 3 个终止密码子中的一个。接着一种释放因子结合到核糖体上，终止了翻译并释放完整的多肽。

真核生物与原核生物的核糖体密切相关，尽管它们的 rRNA 和蛋白质成分的数量与大小方面存在差别。rRNA 在翻译中发挥主要作用，它决定了核糖体的整体结构，形成了 tRNA 结合位点，并将 tRNA 与 mRNA 中的密码子匹配，还提供了肽酰转移酶活性，在翻译时将氨基酸连接起来。

在蛋白质合成的最后一步，两种不同的分子伴侣指导了多肽链的折叠。这些称作 hsp60 和 hsp70 的伴侣蛋白，识别蛋白质上暴露出来的疏水区域，并帮助防止蛋白质聚集，在其他情况下，这种聚集会与新合成的蛋白质折叠成正确的三维构象竞争。这种蛋白质折叠过程也必须与高度复杂的质量控制机制竞争，该机制破坏具有异常暴露疏水区域的蛋白质。在这种情况下，通过泛素连接酶将泛素共价结合到错误折叠的蛋白质上，结果形成的多泛素链由蛋白酶体上的帽子识别，从而将整个蛋白质转移至蛋白酶体内部用于蛋白酶解降解。存在一种密切相关的蛋白酶解机制，基于由泛素连接酶识别的特殊降解信号，用来决定许多正常折叠蛋白的生命周期。通过这种机制，将被选定的正常蛋白质从细胞清除，作为对特定信号的应答。

RNA 世界与生命的起源

为了完全理解现今活细胞内部进行的过程，我们需要考虑它们是怎样在进化中出现的。所有这类问题中最基本的就是遗传信息的表达，在今天它需要特别复杂的装置，并经过一个 RNA 中间物后，从 DNA 到达蛋白质。这种机制怎样产生的呢？一种观点认为在现代细胞出现以前，地球上存在一个 RNA 世界（图 6-91）。根据这种假说，在原始细胞中，RNA 不仅储存了遗传信息，而且催化了化学反应。只是在进化的后期，DNA 才取代 RNA 作为遗传物质，而蛋白质也成为细胞主要的催化剂和结构组分。如果这种观点是正确的，那么从 RNA 世界进行的转变尚未完成，因为本章我们已经看到，RNA 仍然催化现代细胞中多种基本反应，因而可将其视为早期世界的分子化石。

这一部分概括了一些支持 RNA 世界假说的观点。我们会看到，对于现今细胞许多更令人奇怪的特征，例如，核糖体和前体 mRNA 的剪接装置，如果把它们看成是 RNA 世界中主宰细胞代谢的由 RNA 介导的相互作用形成的互杂网络的衍生物，就非常容易解释了。我们也讨论 DNA 可能是怎样接替作为遗传物质的，遗传密码是怎样出现的，以及蛋白质是怎样

取代 RNA 行使现今细胞中大量的生物化学催化作用的。

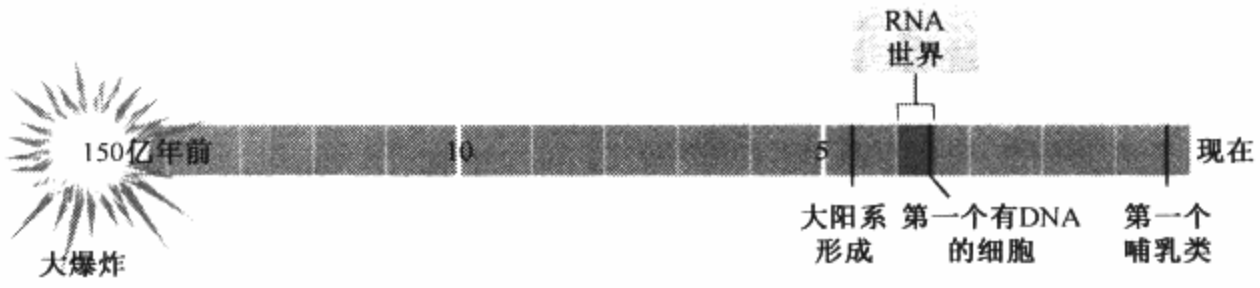


图 6-91　宇宙的时间线，表明了生物系统早期存在着一个 RNA 世界

生命需要自我催化作用（自催化）

曾有人提出地球上第一个“生物”分子是由基于金属的催化作用在矿物的结晶表面形成的。从理论上讲，一种复杂的分子合成及降解体系（代谢）早在第一个细胞出现以前，是可能在这些表面上存在的。但是生命需要具有一种关键特性的分子：能够催化那些直接或间接生成与它们自身相似的更多分子的反应。拥有这种特殊的自我促进功能的催化剂能利用原材料对自身进行增殖，因而改变这些相同原材料的去向，不再生成其他物质。但是在早期细胞中什么样的分子有这种自我催化功能呢？在现今细胞中，功能最多样的自我催化剂就是多肽，它们由许多氨基酸组成，具有化学上不同的侧链，因此，能够采用充满活性化学基团的不同三维构象。但是，虽然多肽作为催化剂时功能多种多样，还没有已知的途径，使一种分子能够直接通过指令形成另外的序列完全相同的分子来对自身进行增殖。

多聚核苷酸不仅能储存信息，还能够催化化学反应

多聚核苷酸拥有一种特性，与多肽的特性形成对比：它们直接指导自身序列的形成。这种能力依赖于核苷酸亚单位的互补碱基配对，这种配对使一条多聚核苷酸能用作形成另一条的模板。我们在本章以及前一章已经看到，这种互补模板机制是现今细胞 DNA 复制和转录的中心所在。

但是由这种互补模板机制进行有效的合成，需要催化剂来促进多聚化反应：没有催化剂，多聚体的形成就很慢，易出现错误，且效率不高。现在基于模板的核苷酸多聚化由蛋白酶迅速地催化，如 DNA 聚合酶和 RNA 聚合酶。但是在具有适当催化特异性的蛋白质存在之前是怎样进行催化的呢？1982 年首先得到了其中一个答案，那时发现 RNA 本身能作为催化剂发挥作用。本章我们已经看到，RNA 分子是在核糖体上进行的肽酰转移酶反应的催化剂。RNA 分子既作为信息载体，又作为催化剂的独特潜能构成了 RNA 世界假说的基础。

因此 RNA 具备了能够催化自身合成的分子所需要的全部特性（图 6-92）。尽管 RNA 自我复制系统在自然界还未发现，但是科学家们充满希望，认为能在实验室构建出来。虽然这个论证不能证明地球

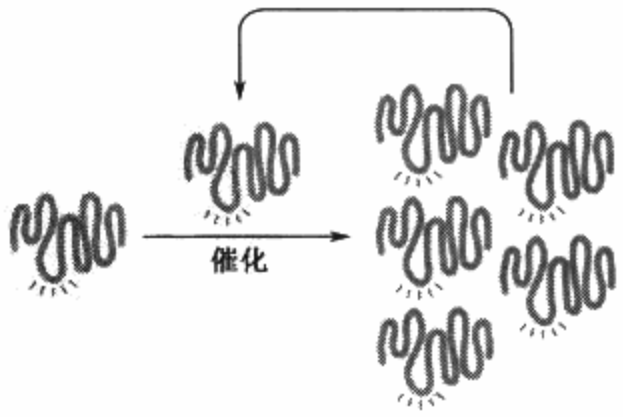


图 6-92　能催化自身合成的 RNA 分子
这个假设的过程要求催化合成具有互补核苷酸序列的另一条 RNA 链和利用这条 RNA 分子作为模板合成许多具有最初序列的 RNA 分子，红色光线表示这个假设 RNA 酶的活性位点

上这种自我复制的 RNA 分子在生命起源中是必需的，但是它将肯定地表明这种情况是有可能发生的。

在 RNA 世界之前可能存在一个前 RNA 世界

尽管 RNA 看起来很适合作为形成一套自我复制的生物化学催化剂的基础，但是 RNA 不太可能是第一种能做到这一点的分子。从纯化学角度来看，很难想像最初通过纯粹的非酶催化方式，能合成多长的分子。一方面，没有酶的催化很难合成 RNA 前体，核糖核苷酸。此外，RNA 的合成要求在面临一组竞争的反应（包括分解反应、2′→5′连接和 5′→5′连接等）时能形成一系列很长的 3′→5′磷酸二酯连接。考虑到这些问题，有人提出第一个既储存信息又具有催化活性的分子可能是类似 RNA 但化学上更简单的多聚物（图 6-93）。在现今细胞中我们没有发现这些化合物的残留，这些化合物也没有留下化石记录。然而这些 RNA 类似多聚物的相对简单性，使它们比 RNA 自身更适合作为地球上第一个既有信息储存能力，又有催化能力的生物多聚体。

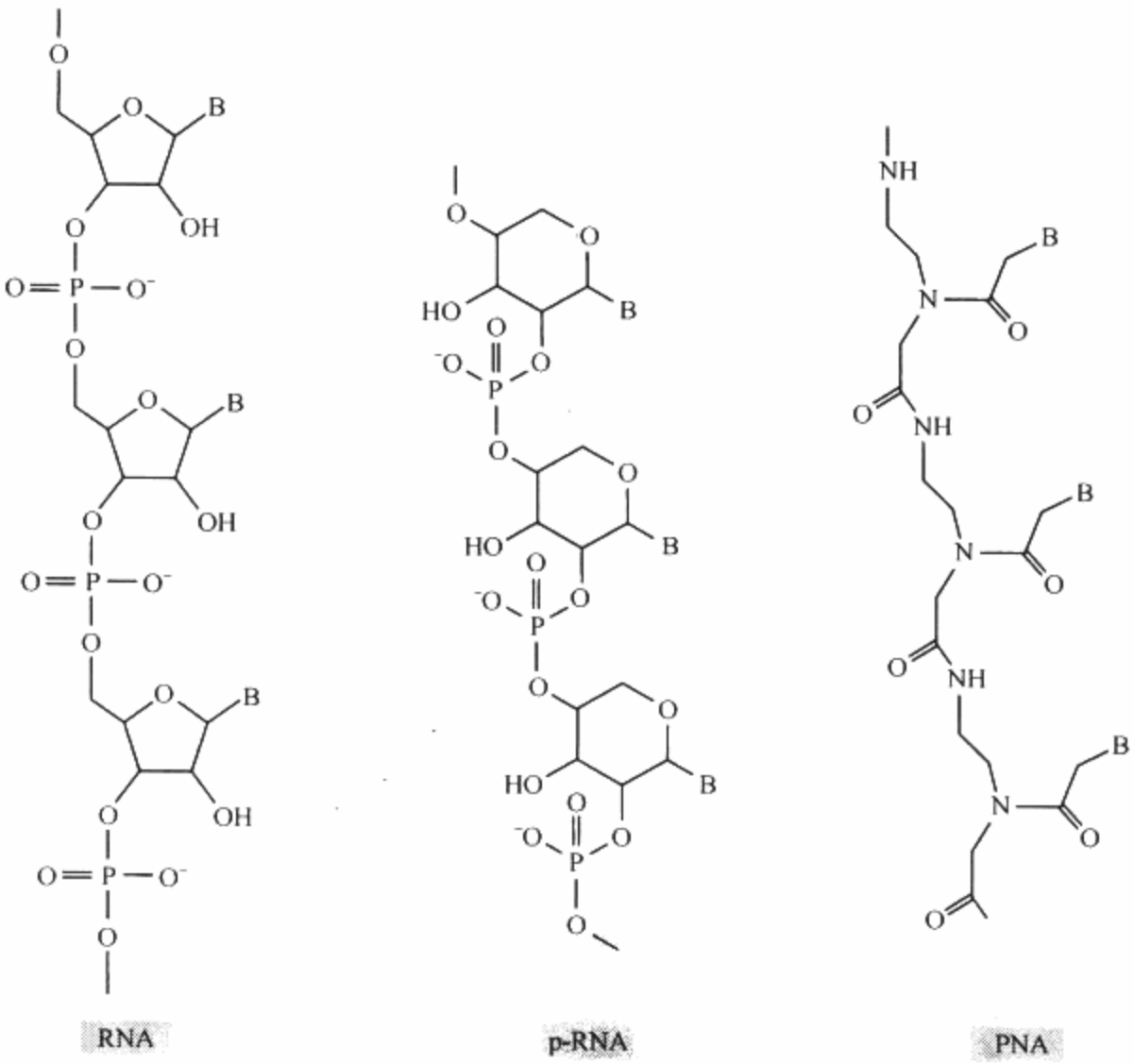


图 6-93　RNA 结构和两个相关的携带信息的多聚物

在每个例子中，B 表示嘌呤和嘧啶碱基的位置。多聚物 p-RNA（吡喃糖基 RNA）是一种核糖的呋喃形式（五元环）已被此吡喃形式（六元环）取代的 RNA。在 PNA（肽核酸）中，RNA 的核糖磷酸骨架由蛋白质中存在的肽骨架取代，与 RNA 一样，p-RNA 和 PNA 都能通过互补碱基配对形成双螺旋，因而从理论上讲每个可能作为模板用于自身合成（图 6-92）

在利用这些更加简单的复合物之一既作为模板，又作为催化剂来合成 RNA 的过程中，可能发生了前 RNA 世界向 RNA 世界的转变。表明这些更加简单的形式（PNA，如图 6-93

所示）能够作为模板合成互补 RNA 分子的实验研究支持了这个体系的合理性，因为两种分子碱基的整体几何特征相似。或许，前 RNA 多聚物也催化从简单的分子合成核糖核苷酸。一旦合成出第一个 RNA 分子，它们就可能逐渐多样化，并取代了起初由前 RNA 多聚物执行的功能，最终导致出现所推测的 RNA 世界。

单链 RNA 分子能折叠成高度复杂的结构

我们已经看到在同一条链的核苷酸之间能形成互补碱基配对和其他类型的氢键，使 RNA 分子以一种独特的方式折叠起来，这种方式由其核苷酸序列决定（图 6-6、图 6-52 和图 6-67）。对许多 RNA 结构进行比较后揭示了保守的模式，也就是一次又一次用作更大结构一部分的短结构元件。这些 RNA 二级结构模式中有些在图 6-94 中加以说明。此外，图 6-95 示出了一些更加复杂且更远距离相互作用的几个常见例子，后者通常称作 RNA 四级相互作用。

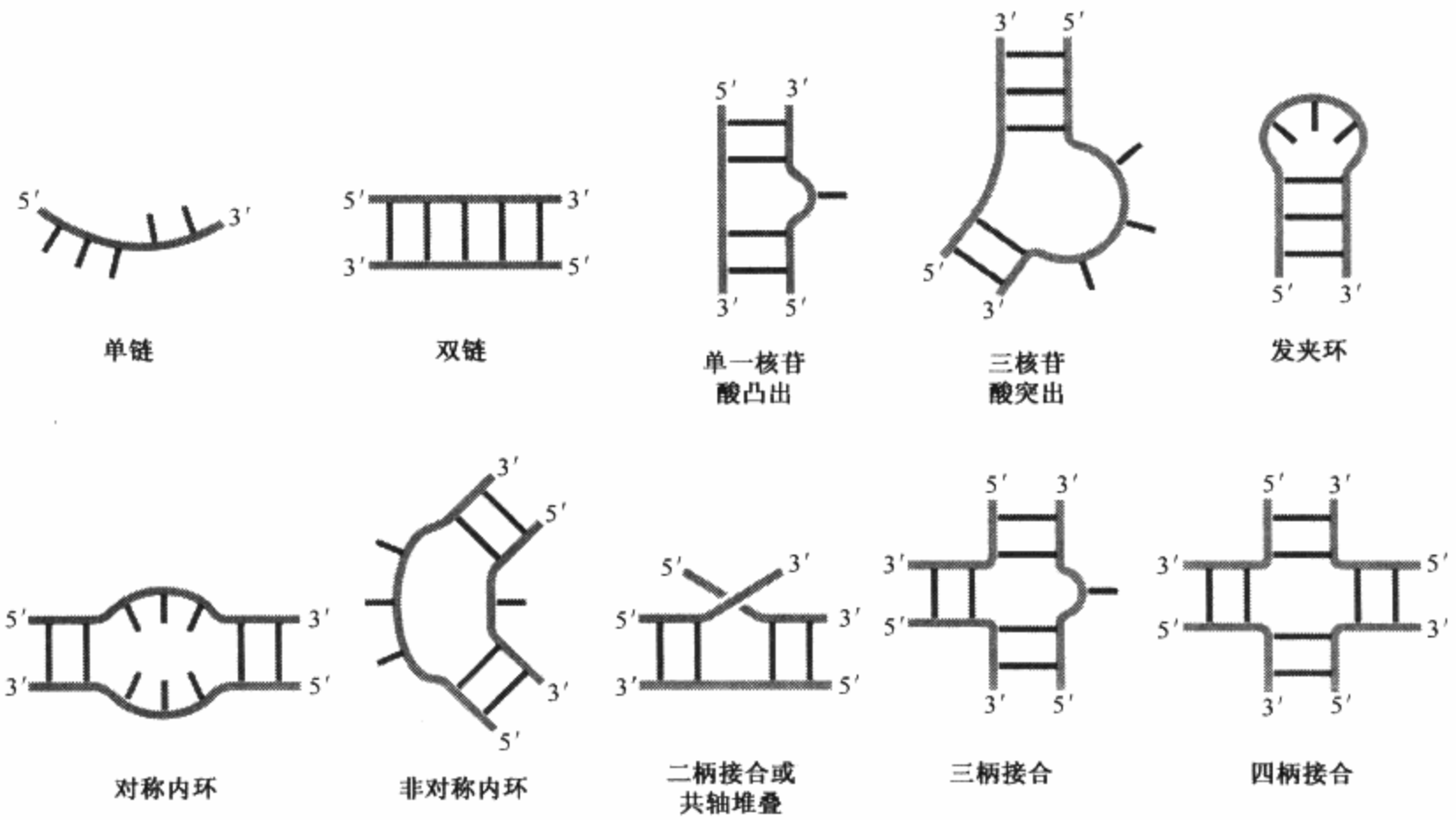


图 6-94　RNA 二级结构常见元件

RNA 双螺旋部分中红色的“梯级”表示常规互补碱基配对相互作用

蛋白质催化剂需要一个具有独特轮廓和化学特性的表面，在其上面给定的一组底物能够发生反应（讨论见第 3 章）。具有适当折叠形状的 RNA 分子能够作为酶，以恰恰相同的方式发挥作用（图 6-96）。与某些蛋白质类似，这些核酶中有许多通过在它们的活性位点定位金属离子发挥功能，这个特征使它们具备更加广泛的催化活性，而不只是因有限的多聚核苷酸链化学基团而具有的活性。

然而，现今细胞中存在相当少的有酶活性的 RNA，我们有关 RNA 世界的推理，大多来自于在实验室合成大量具有随机核苷酸序列的 RNA 分子的实验。那些具有实验规定特性的稀有的 RNA 分子被挑选出来进行研究（图 6-97）。这类实验制造出能催化大量不同生物化学

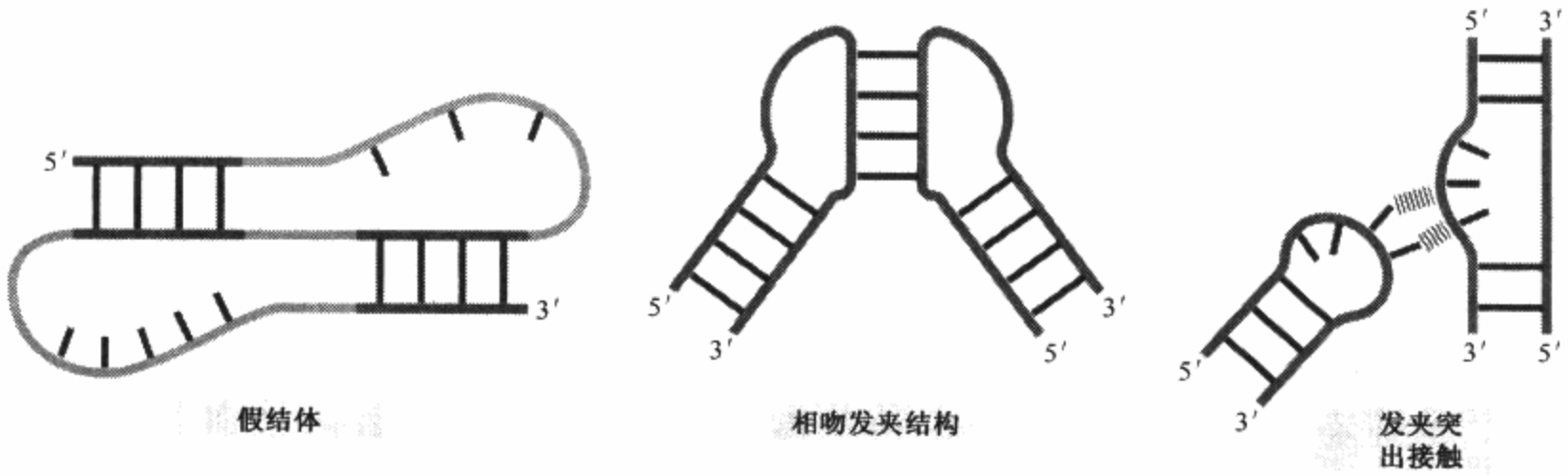


图 6-95　RNA 三级相互作用实例

这些相互作用中有些能够将同一 RNA 分子相距甚远的部分结合起来，或者将两条单独的 RNA 分子并到一起

反应的 RNA（表 6-4），并表明蛋白酶和核酶之间主要的差别在于它们的最大反应速度，而不是它们所能催化反应的多样性。

与蛋白质类似，RNA 在小分子，或其他 RNA 的诱导下能发生别构构象变化。一种人工合成的核酶能以两种完全不同的构象存在，每一种都有不同的催化活性（图 6-98）。此外，核糖体中 rRNA 自身的结构和功能已经表明 RNA 是功能极其多样的分子。因而容易想像到 RNA 世界可以具有很高水平的生物化学复杂性。

自我复制的分子经历自然选择

多聚核苷酸的三维折叠结构影响它的稳定性，它对其他分子的作用，以及它复制的能力。因此，某些多聚核苷酸在任何自我复制的混合物中都将特别成功。由于错误不可避免地出现在任何拷贝过程中，所以经过长期一段时间以后就会生成这些核苷酸的变体序列。

某种催化活性在生命早期进化过程中可能是最重要的。尤其考虑到帮助催化模板化聚合过程的某个 RNA 分子，该聚合过程将任何给定的 RNA 分子作为模板（这种核酶活性已在体外直接证明，虽然是处于一种只能合成中等长度 RNA 的初级形式）。这样的分子通过作用于自身拷贝，能够进行复制。同时它能促进邻近的其他类型 RNA 分子的复制（图 6-99）。如果这些相邻 RNA 中有些具有催化活性，帮助 RNA 以其他方式（如催化核糖核苷酸的合成）存活下来，那么各自都专门具有不同活性的一组不同类型的 RNA 分子，可能演变成一个协同系统，以极高的效率进行复制。

导致形成有效的自我复制系统的关键事件之一，必定是单个区室的发展。例如，对于一组彼此有利的 RNA（如图 6-99 中的那些），只有当所有这些 RNA 都保留在专门负责模板多聚化的 RNA 附近时，才能对自身进行复制。此外，如果这些 RNA 能在大群其他 RNA 分子中自由扩散，它们就可能加入其他复制系统，这些系统继而就可能与最初的 RNA 系统竞争原材料。根据 RNA 生成的自我复制系统的质量是不能有效地选择出一组 RNA 分子的，直到某些区室形式经过演化，将它们包含进去，并因而使它们只由生成它们的 RNA 所接触。早期原始的区室化可能就是在表面或颗粒上简单的吸附。

对于更加复杂的包容类型的需要很容易由一类小分子实现，它们具有简单的物理化学特性——两亲性，即由疏水部分（水不溶性）和亲水部分（水溶性）组成。将这样的分子放到

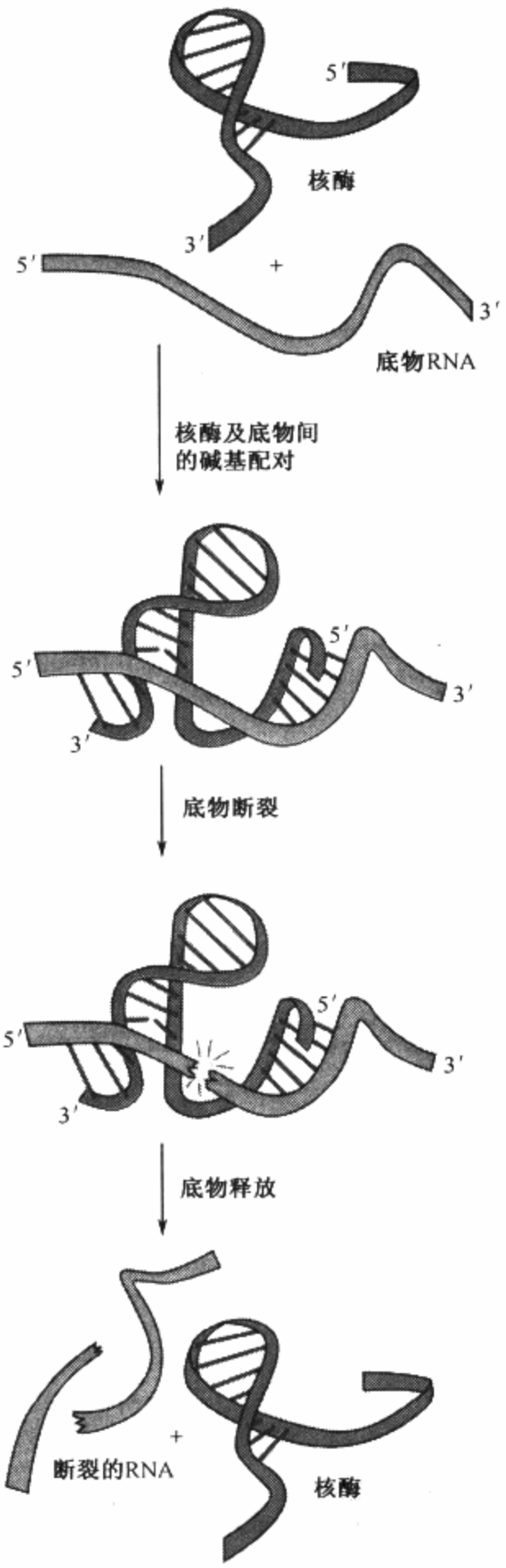


图 6-96　核酶

这个简单的 RNA 分子催化另一条 RNA 分子在特殊位点的切割。该核酶发现深埋在感染植物的更大的 RNA 基因组——类病毒中。自然界中含有该核酶的相同 RNA 分子在相距甚远的位置上发生的切割，是类病毒基因组复制中的一个步骤。虽然未在图中显示，但该反应需要一分子定位于活性位点的 Mg（改编自 T. R. Cech and O. C. Uhlenbeck，*Nature* 373：39～40，1994）

水中时，它们会聚集起来，使疏水部分尽可能彼此接触，亲水部分则与水接触。具有适当形状的两亲分子能自发聚集起来形成双层，产生封闭的小泡，将内部含水成分与外部介质隔

表 6-4　一些可以被核酶催化的生化反应

活性	核酶
蛋白质合成中的肽键形成	核糖体 RNA
RNA 断裂 RNA 连结	自我剪切 RNA，及体外选择 RNA
DNA 断裂	自我剪切 RNA
RNA 剪切	自我剪切 RNA，也许是剪切体的 RNA
RNA 聚合	体外选择的 RNA
RNA，DNA 磷酸化	体外选择的 RNA
RNA 氨基酰化	体外选择的 RNA
RNA 烷	体外选择的 RNA
酰胺键形成	体外选择的 RNA
酰胺键断裂	体外选择的 RNA
糖苷键形成	体外选择的 RNA
卟啉金属化作用	体外选择的 RNA

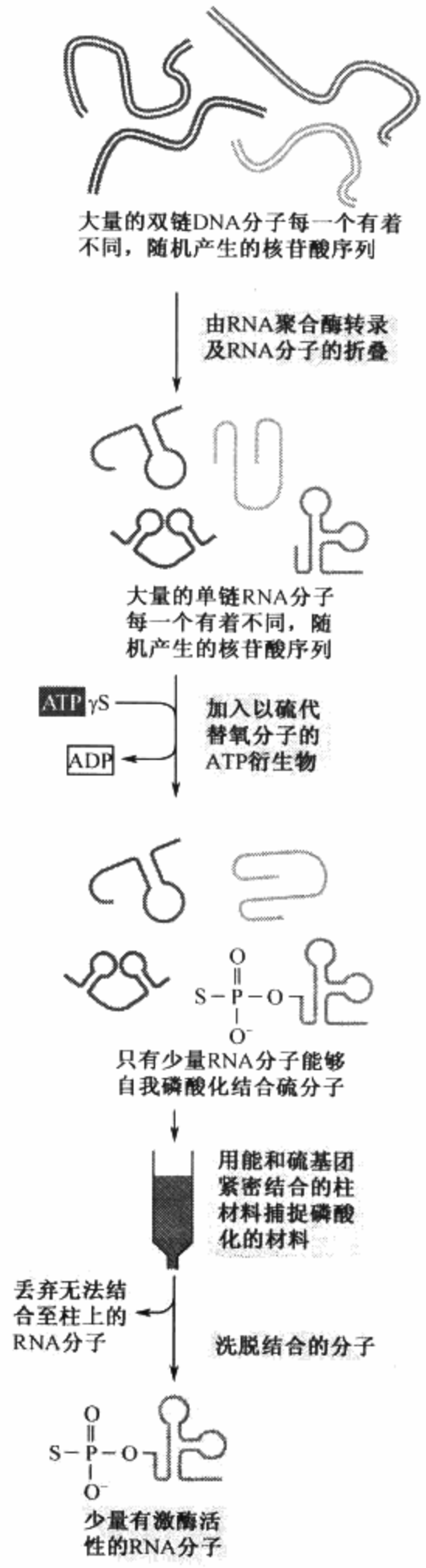


图 6-97　合成核酶的体外筛选

从实验室合成的巨大核酸分子库开始，可以分离出那些获得指定催化活性的稀有 RNA 分子并加以研究。虽然图中显示了特定的一个例子（自我磷酸化的核酶），但这个流程的变通形式已用于产生表 6-4 中所列的许多核酶。在自我磷酸化步骤中，对 RNA 分子进行足够的稀释，以防止与其他 RNA 分子的“交叉”磷酸化。实际上，必需重复此过程以筛选出具有催化活性的非常稀有的 RNA 分子。因此最初从柱中洗脱的物质转变为 DNA，并进行多倍扩增（利用第 8 章描述的反转录酶和 PCR），再转录成 RNA，进行重复筛选（摘自 J. R. Lorsch and J. W. Szostak，*Nature* 371：31～36，1994）

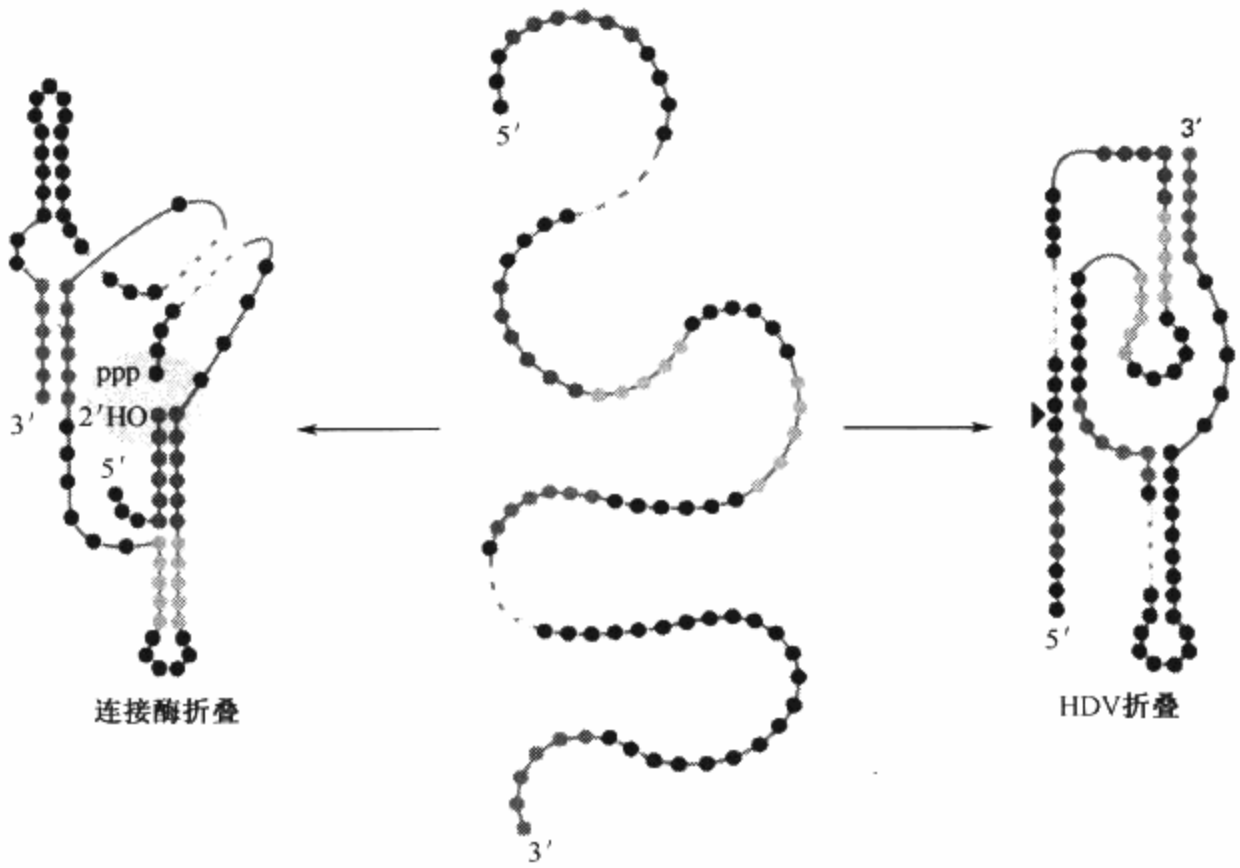


图 6-98　折叠成两种不同核酶的 RNA 分子

这条在实验室合成的长 88 个核苷酸的 RNA，能折叠成催化自我连接反应（左边）或自我切割反应（右边）的核酶。连接反应形成一个 2′→5′磷酸二酯键，并释放出焦磷酸。此反应将通过实验方法引入到该 RNA 分子中的缺口封闭（灰色阴影）。在 HDV 折叠催化的反应中，RNA 在同一位置切割，用箭头表示。这种切割类似于 HDV，一种 B 型肝炎卫星病毒，生命周期中所使用的切割，因而如此命名。每个核苷酸用彩色点表示，所使用的颜色只是为了说明清楚两种不同的折叠模式。折叠的结构说明了两种核酶中具有碱基配对区域的二级结构，这些区域用紧密相对的彩色点表示。注意这两种核酶没有共同的二级结构（摘自 E. A. Schulters and D. P. Bartel, *Science* 289: 448～452, 2000）

离开（图 6-100）。这种现象可以在试管中演示，只需将磷脂和水混合，在合适的条件下，将会形成小泡。所有现今的细胞都由这种构型的两亲分子（主要是磷脂）组成的细胞质膜包被起来；我们将在第 10 章详细讨论这些分子。

可能第一个膜包被的细胞通过一些两亲分子自发组装而成，将自发进行复制的 RNA（或前 RNA）和其他分子的混合物包围了起来。还不清楚在生物催化剂进化的哪个点上首先出现这种情况。无论如何，一旦 RNA 分子密封到封闭的膜内，它们就能最早开始进化成遗传指令的携带者：不仅基于它们的结构进行选择，而且根据它们对同一区室内其他分子的影响进行选择。现在 RNA 分子的核苷酸序列能以符合单个活细胞的特征进行表达。

蛋白质合成是怎样进化的

现今细胞中蛋白质合成的分子过程似乎复杂到无法摆脱的地步。虽然我们了解了它们中的大多数，但是它们却不具备像 DNA 转录、DNA 修复、DNA 复制那样在概念上的意义。想像蛋白质合成是如何进化的特别困难，因为这个过程现在由蛋白质和 RNA 分子形成的复杂连锁系统完成；显然蛋白质是直到早期反应装置已经存在之后才出现的。尽管我们只能推测蛋白质合成和遗传密码的起源，但是许多实验方法已经提供了可能发生过的情节。

体外对前面图 6-97 中概述的 RNA 类型进行选择实验，合成出了能紧密结合氨基酸的

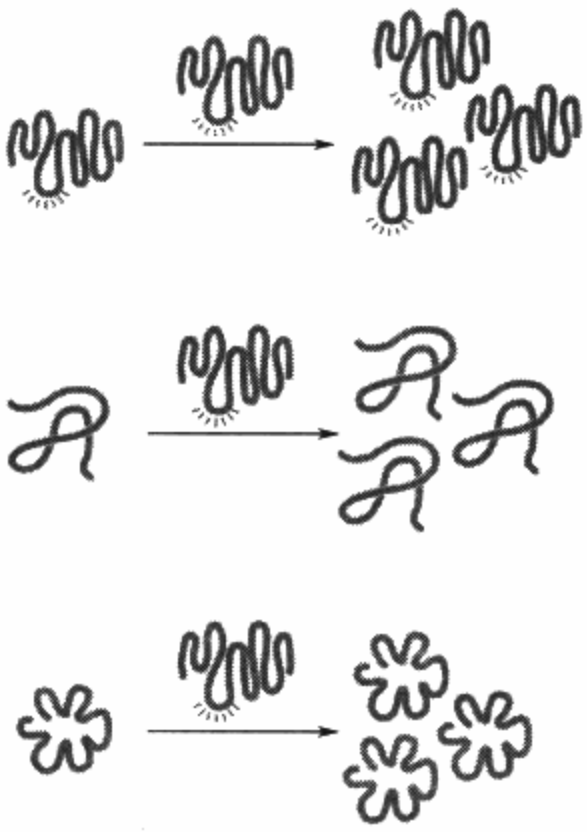


图 6-99　互助 RNA 分子家族，彼此催化对方的合成

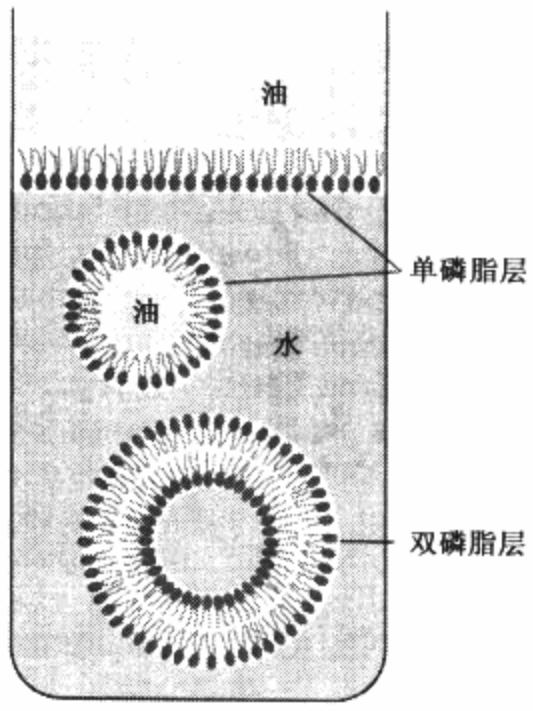


图 6-100　由磷脂形成膜

由于这些分子具有亲水的头部和亲脂的尾部，所以它们在油/水交界面的排列是头部朝向水面而尾部朝向油面。在水中它们彼此结合形成封闭的双层小泡，其中亲脂尾部彼此接触而亲水头部暴露在水中

RNA 分子。这些 RNA 的核苷酸序列中通常含有它们识别的氨基酸的密码子，且它们的频率高得不呈比例。例如，能够选择性地结合 Arg 的 RNA 分子中 Arg 密码子占优势，而结合 Tyr 的 RNA 分子中 Tyr 密码子占优势。这种相关性并非对于所有氨基酸都是符合的，它的解释也众说不一，但是它提出了一种可能性，就是有限的遗传密码可能因为氨基酸直接与特定 RNA 序列结合而产生，其中 RNA 作为原始的模板，指导多个不同氨基酸的非随机聚合。在前面描述的 RNA 世界里，任何帮助指导合成一条有用多肽的 RNA，都将在为生存而竞争的进化中拥有巨大的优势。

在现今细胞中，tRNA 适配器用于将氨基酸和密码子匹配，蛋白质则催化了 tRNA 氨酰化反应。然而，在实验室合成的核酶能进行特定 tRNA 氨酰化反应，所以有可能类似 tRNA 的适配器已经在 RNA 世界出现了。这样的发展可能已经使“mRNA”序列与氨基酸的匹配更加有效，并且它可能使能够用于基于模板的蛋白质合成的氨基酸数量增加。

最后，早期蛋白质合成形式的效率会因为肽键形成的催化作用而显著提高。对于这种进化发展不会有任何概念上的问题，因为我们已经看到，这种反应由现今细胞中的 rRNA 催化。可以想像存在一个原始的肽基转移核酶，它经过多年以后变得更大，并且获得了将带电荷的 tRNA 精确定位在 RNA 模板上的能力，最终导致形成现代核糖体。一旦蛋白质合成进化出来了，那么向蛋白质世界的转变就可以向前进行，由于拥有 20 个而不是 4 个不同的亚单位，蛋白质具有更大的功能多样性，所以最终接管了大部分的催化和结构任务。

现今所有细胞利用 DNA 作为它们的遗传物质

比起现今即便是最简单的细胞，RNA 世界的细胞也要简单得多，且增殖自身效率要低得多，因为 RNA 分子的催化作用比蛋白质的催化作用效率更低。它们可能只不过由一层简单的外膜包围一组自我复制的分子，以及其他一些需要为它们的复制提供原料和能量的分子而形成。即使上面概括的有关 RNA 的进化推测是正确的，那么这些早期细胞也将从根本上与我们今天所了解的细胞不一样，因为它们将遗传信息储存在 RNA 而不是 DNA 中(图6-101)。

有关 RNA 在进化中先于 DNA 出现的证据也能从它们之间的差别中找到。与葡萄糖和其他简单的碳水化合物类似，核糖能够由甲醛（HCHO）合成得到。甲醛是一种简单的化学物质，容易在模拟原始地球条件的实验室中合成。脱氧核糖则更难获得，在现今细胞中，它的生成是在一种蛋白质催化的反应中由核糖得到的，表明细胞中核糖先于脱氧核糖出现。DNA 可能在后来的发生的情节中才出现，而且后来证明比 RNA 更适合作为遗传信息的永久储存库，尤其是位于 DNA 链中的糖磷酸骨架使它在化学上比 RNA 链更加稳定，这样就可以维持更长的 DNA，而不至于断裂。

DNA 和 RNA 之间的其他差别（DNA 的双螺旋结构以及利用了胸腺嘧啶而不是尿嘧啶）进一步增强了 DNA 的稳定性，这是通过使许多在 DNA 分子上发生的不可避免的事故更容易修复而实现的，这些在第 5 章已详细讨论过。

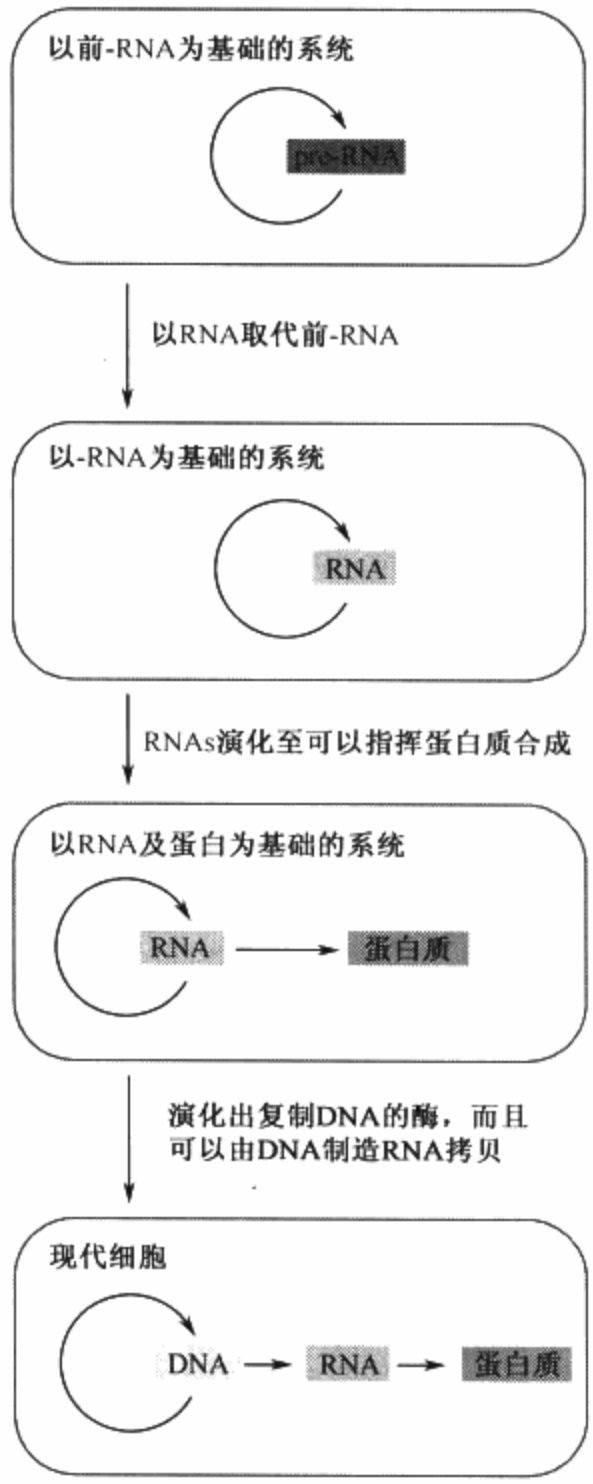


图 6-101　有关 RNA 在进化过程中先于 DNA 和蛋白质存在的假说

在最早的细胞中，前 RNA 分子可能综合了遗传、结构、催化的功能，并且这些功能可能已经逐渐由 RNA 取代。在现今细胞中，DNA 是遗传信息的储备库，蛋白质在细胞中行使大多数催化功能。今天 RNA 主要作为蛋白质合成过程中的中间角色，尽管它仍然是催化若干关键反应的酶

小结

从我们有关现今生物及其所含分子的知识来看，对于生命系统很重要的直接自我催化机制的发展，似乎有可能开始于能够催化自身复制的分子家族的进化，一群协同作用的 RNA 催化剂可能随时间演化出指导多肽合成的能力。DNA 有可能是后来加入的：由于其他蛋白质催化剂的累积，使更有效且更复杂的细胞得以进化，DNA 双螺旋作为更加稳定的分子来储存这些细胞所需的数量增加的遗传信息，从而取代了 RNA。

参 考 文 献

一般文献

Gesteland RF, Cech TR & Atkins JF (eds) (1999) The RNA World, 2nd edn. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.

Hartwell L, Hood L, Goldberg ML et al. (2000) Genetics: from Genes to Genomes. Boston: McGraw Hill.

Lewin B (2000) Genes VII. Oxford: Oxford University Press.

Lodish H, Berk A, Zipursky SL et al. (2000) Molecular Cell Biology, 4th edn. New York: WH Freeman.

Stent GS (1971) Molecular Genetics: An Introductory Narrative. San Francisco: WH Freeman.

Stryer L (1995) Biochemistry, 4th edn. New York: WH Freeman.

Watson JD, Hopkins NH, Roberts JW et al. (1987) Molecular Biology of the Gene, 4th edn. Menlo Park, CA: Benjamin/Cummings.

From DNA to RNA

Berget SM, Moore C & Sharp PA (1977) Spliced segments at the 5′ terminus of adenovirus 2 late mRNA. *Proc. Natl. Acad. Sci. USA* 74, 3171–3175.

Black DL (2000) Protein diversity from alternative splicing: a challenge for bioinformatics and post-genome biology. *Cell* 103, 367–370.

Brenner S, Jacob F & Meselson M (1961) An unstable intermediate carrying information from genes to ribosomes for protein synthesis. *Nature* 190, 576–581.

Cech TR (1990) Nobel lecture. Self-splicing and enzymatic activity of an intervening sequence RNA from Tetrahymena. *Biosci. Rep.* 10, 239–261.

Chow LT, Gelinas RE, Broker TR et al. (1977) An amazing sequence arrangement at the 5′ ends of adenovirus 2 messenger RNA. *Cell* 12, 1–8.

Conaway JW, Shilatifard A, Dvir A & Conaway RC (2000) Control of elongation by RNA polymerase II. *Trends Biochem. Sci.* 25, 375–380.

Cramer P, Bushnell DA, Fu J et al. (2000) Architecture of RNA polymerase II and implications for the transcription mechanism. *Science* 288, 640–649.

Crick F (1979) Split genes and RNA splicing. *Science* 204, 264–271.

Daneholt B (1997) A look at messenger RNP moving through the nuclear pore. *Cell* 88, 585–588.

Darnell JE, Jr. (1985) RNA. *Sci. Am.* 253(4), 68–78.

Dvir A, Conaway JW & Conaway RC (2001) Mechanism of transcription initiation and promoter escape by RNA polymerase II. *Curr. Opin. Genet. Dev.* 11, 209–214.

Ebright RH (2000) RNA polymerase: structural similarities between bacterial RNA polymerase and eukaryotic RNA polymerase II. *J. Mol. Biol.* 304, 687–698.

Eddy SR (1999) Noncoding RNA genes. *Curr. Opin. Genet. Dev.* 9, 695–699.

Green MR (2000) TBP-associated factors (TAFIIs): multiple, selective transcriptional mediators in common complexes. *Trends Biochem. Sci.* 25, 59–63.

Harley CB & Reynolds RP (1987) Analysis of E. coli promoter sequences. *Nucleic Acids Res.* 15, 2343–2361.

Hirose Y & Manley JL (2000) RNA polymerase II and the integration of nuclear events. *Genes Dev.* 14, 1415–1429.

Kadonaga JT (1998) Eukaryotic transcription: an interlaced network of transcription factors and chromatin-modifying machines. *Cell* 92, 307–313.

Lewis JD & Tollervey D (2000) Like attracts like: getting RNA processing together in the nucleus. *Science* 288, 1385–1389.

Lisser S & Margalit H (1993) Compilation of E. coli mRNA promoter sequences. *Nucleic Acids Res.* 21, 1507–1516.

Minvielle-Sebastia L & Keller W (1999) mRNA polyadenylation and its coupling to other RNA processing reactions and to transcription. *Curr. Opin. Cell Biol.* 11, 352–357.

Mooney RA & Landick R (1999) RNA polymerase unveiled. *Cell* 98, 687–690.

Olson MO, Dundr M & Szebeni A (2000) The nucleolus: an old factory with unexpected capabilities. *Trends Cell Biol.* 10, 189–196.

Proudfoot N (2000) Connecting transcription to messenger RNA processing. *Trends Biochem. Sci.* 25, 290–293.

Reed R (2000) Mechanisms of fidelity in pre-mRNA splicing. *Curr. Opin. Cell Biol.* 12, 340–345.

Roeder RG (1996) The role of general initiation factors in transcription by RNA polymerase II. *Trends Biochem. Sci.* 21, 327–335.

Shatkin AJ & Manley JL (2000) The ends of the affair: capping and polyadenylation. *Nat. Struct. Biol.* 7, 838–842.

Smith CW & Valcarcel J (2000) Alternative pre-mRNA splicing: the logic of combinatorial control. *Trends Biochem. Sci.* 25, 381–388.

Staley JP & Guthrie C (1998) Mechanical devices of the spliceosome: motors, clocks, springs, and things. *Cell* 92, 315–326.

Tarn WY & Steitz JA (1997) Pre-mRNA splicing: the discovery of a new spliceosome doubles the challenge. *Trends Biochem. Sci.* 22, 132–137.

von Hippel PH (1998) An integrated model of the transcription complex in elongation, termination, and editing. *Science* 281, 660–665.

From RNA to Protein

Abelson J, Trotta CR & Li H (1998) tRNA splicing. *J. Biol. Chem.* 273, 12685–12688.

Anfinsen CB (1973) Principles that govern the folding of protein chains. *Science* 181, 223–230.

Cohen FE (1999) Protein misfolding and prion diseases. *J. Mol. Biol.* 293, 313–320.

Crick FHC (1966) The genetic code: III. *Sci. Am.* 215(4), 55–62.

Fedorov AN & Baldwin TO (1997) Cotranslational protein folding. *J. Biol. Chem.* 272, 32715–32718.

Frank J (2000) The ribosome—a macromolecular machine par excellence. *Chem. Biol.* 7, R133–141.

Green R (2000) Ribosomal translocation: EF-G turns the crank. *Curr. Biol.* 10, R369–373.

Hartl FU (1996) Molecular chaperones in cellular protein folding. *Nature* 381, 571–580.

Hershko A, Ciechanover A & Varshavsky A (2000) The ubiquitin system. *Nat. Med.* 6, 1073–1081.

Ibba M & Soll D (2000) Aminoacyl-tRNA synthesis. *Annu. Rev. Biochem.* 69, 617–650.

Kozak M (1999) Initiation of translation in prokaryotes and eukaryotes. *Gene* 234, 187–208.

Nissen P, Hansen J, Ban N et al. (2000) The structural basis of ribosome activity in peptide bond synthesis. *Science* 289, 920–930.

Nureki O, Vassylyev DG, Tateno M et al. (1998) Enzyme structure with two catalytic sites for double-sieve selection of substrate. *Science* 280, 578–582.

Prusiner SB (1998) Nobel lecture. Prions. *Proc. Natl. Acad. Sci. USA* 95, 13363–13383.

Rich A & Kim SH (1978) The three-dimensional structure of transfer RNA. *Sci. Am.* 238(1), 52–62.

Sachs AB & Varani G (2000) Eukaryotic translation initiation: there are (at least) two side to every story. *Nat. Struct. Biol.* 7, 356–361.

The Genetic Code. (1966) Cold Spring Harbor Symposium on Quantitative Biology, vol XXXI. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.

Turner GC & Varshavsky A (2000) Detecting and measuring cotranslational protein degradation *in vivo*. *Science* 289, 2117–2120.

Varshavsky A, Turner G, Du F et al. (2000) The ubiquitin system and the N-end rule pathway. *Biol. Chem.* 381, 779–789.

Voges D, Zwickl P & Baumeister W (1999) The 26S proteasome: a molecular machine designed for controlled proteolysis. *Annu. Rev. Biochem.* 68, 1015–1068.

Wilson KS & Noller HF (1998) Molecular movement inside the translational engine. *Cell* 92, 337–349.

Wimberly BT, Brodersen DE, Clemons WM et al. (2000) Structure of the 30S ribosomal subunit. *Nature* 407, 327–339.

Yusupov MM, Yusupova GZ, Baucom A et al. (2001) Crystal structure of

the ribosome at 5.5A resolution. *Science* 292, 883–896.

The RNA World and the Origins of Life

Bartel DP & Unrau PJ (1999) Constructing an RNA world. *Trends Cell Biol.* 9, M9–M13.

Joyce GF (1992) Directed molecular evolution. *Sci. Am.* 267(6), 90–97.

Knight RD & Landweber LF (2000) The early evolution of the genetic code. *Cell* 101, 569–572.

Orgel L (2000) Origin of life. A simpler nucleic acid. *Science* 290, 1306–1307.

Szathmary E (1999) The origin of the genetic code: amino acids as cofactors in an RNA world. *Trends Genet.* 15, 223–229.

7　基因表达的调控

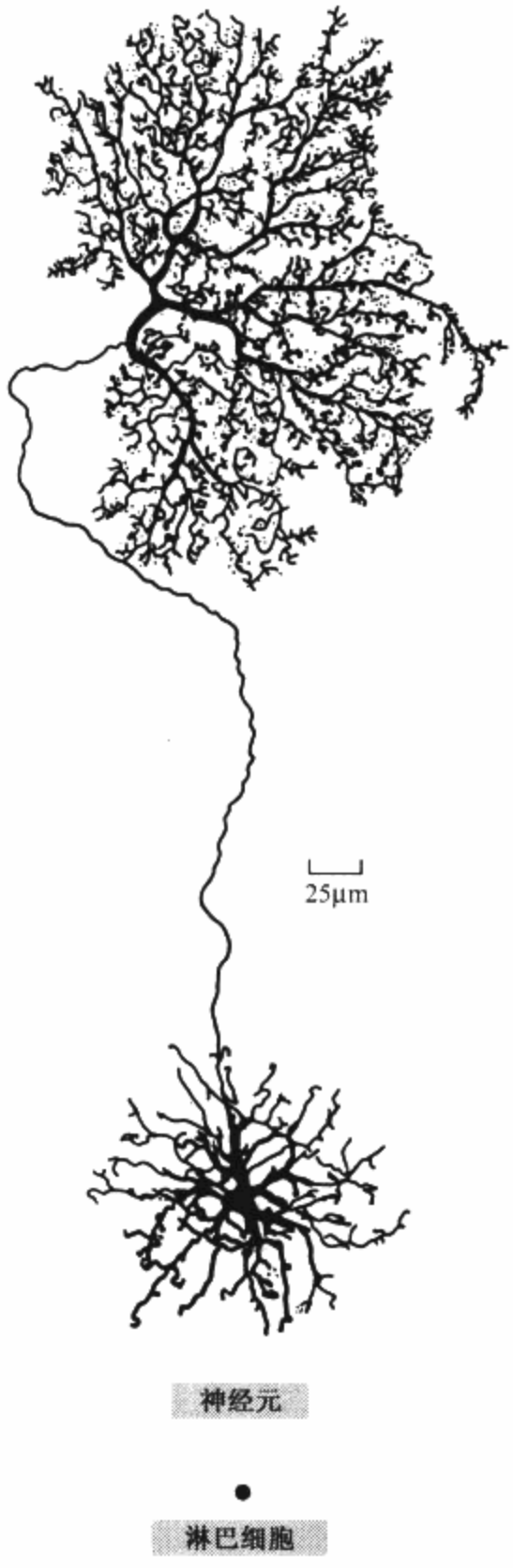


图 7-1　哺乳动物神经元和淋巴细胞

该神经元来自视网膜，它的长分枝使其能够从许多细胞接受电信号并传送这些信号到相邻的细胞中。淋巴细胞是一种白血细胞，参与针对感染的免疫反应，并在体内自由移动。这两种细胞含有相同的基因组，但是它们表达不同的 RNA 和蛋白质[摘自 B. B. Boycott, Essays on the Nervous System (R. Bellairs and E. G. Gray, eds.). Oxford, UK: Clarendon Press, 1974]

基因调控概述
基因调节蛋白中的 DNA 结合基序
基因开关如何工作
产生特化细胞类型的分子遗传学机制
转录后调控
基因组如何进化

每种生物的 DNA 编码了所有构建它的细胞所需的 RNA 和蛋白质分子。可是完整地描述一种生物的 DNA 序列——无论是细菌的几百万个核苷酸，还是人类的几十亿个核苷酸，是不能让我们重建该生物体的，正如一个英语词汇表不能让我们得以重现莎士比亚的戏剧一样。在这两种情况下，问题在于要知道 DNA 序列中的元件或词汇表上的单词是如何使用的。在什么情况下生产每种基因产物，以及一旦生产出来以后它能做些什么？

本章我们讨论这个问题的前半部分，一部分基因在细胞内选择性表达的规则及机制。调控基因表达的机制在许多水平上发挥作用，我们将依次讨论不同的水平。在本章的最后，我们将解释现代基因组以及它们的调节系统是如何在进化过程中形成的。现在从认识多细胞生物基因调控的基本原理开始。

基因调控概述

多细胞生物中不同类型的细胞在结构和功能上显著不同。例如，如果我们将哺乳动物神经元与淋巴细胞进行比较，差别是如此之大以至于难以想像这两种细胞含有相同的基因组（图 7-1）。因为如此，还因为细胞分化往往是不可逆转的，所以生物学家最初怀疑，在细胞分化时基因可能选择性地丢失了。然而现在我们知道了，细胞分化通常依赖于基因表达的变化而不是细胞基因组核苷酸序列的变化。

多细胞生物中不同类型的细胞含有相同的 DNA

多细胞生物中细胞类型彼此互不相同是因为它们合成并积聚一套不同的 RNA 和蛋白质分子。它们通常不需要改变 DNA 序列就能做到这些。细胞分化期间基因组保持不变的证据来自对青蛙进行的一组经典的实验。将一个完全分化的青蛙细胞的细胞核注入一个摘除细胞核的青蛙卵中。供体细胞核能够指导受体卵形成正常的蝌蚪[图7-2(A)]。由于该蝌蚪含有一套完整的分化的细胞，而它们的 DNA 序列来源于最初的供体细胞核，所以分化的供体细胞没有丢失任何重要的 DNA 序列。从不同的植物实验中也得到了相似的结论。实验中将已分

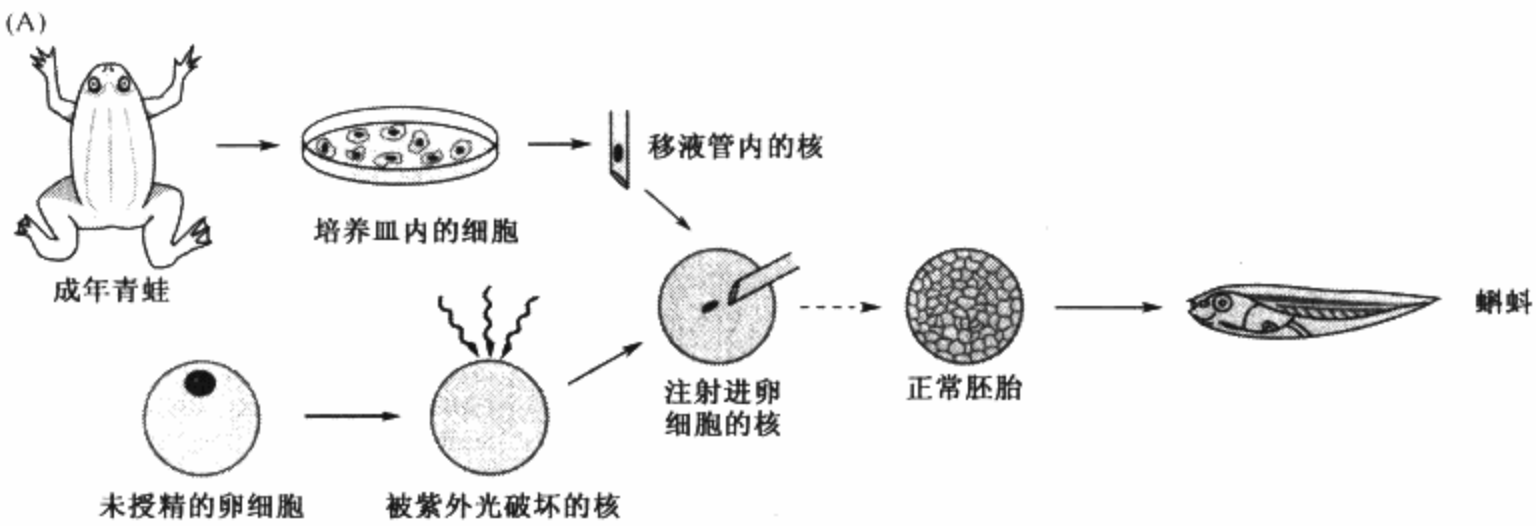


图 7-2　关于分化细胞含有指导形成完整生物体必需的全部遗传指令的证据

(A) 将来自成年青蛙皮肤细胞中的核移植到去核卵细胞中后能发育成整个蝌蚪。断裂箭头表示如果要让移植的基因组有时间适应胚胎环境，就需要进一步的转移程序，即从开始发育的早期胚胎中取出一个核，然后把它放回到另一个去核卵中。(B) 在许多种植物里，分化细胞保留“去分化”的能力，这样单个细胞可以形成后代细胞克隆群，由它们发育成完整的植株。(C) 来自成年母牛的分化细胞引入到来自不同母牛的去核卵中以后能发育成小牛。从同一供体分化细胞产生的不同小牛在遗传上是一样的，因而互为彼此的克隆[(A)改编自 J. B. Gurdon, *Sci. Am.* 219(6):24～35,1968]

化的植物组织块放到培养基中，然后将它们解离成单个细胞。这些细胞中有一些能再生出一株成熟的植株[图7-2(B)]。近来也在哺乳动物包括绵羊、牛、猪、山羊中验证了相同的原理。通过将体细胞核引入去核卵细胞中；将卵细胞植入代孕母体中以后，其中一些卵（称为重建合子）发育成健康的动物[图7-2(C)]。

有关在脊椎动物发育期间并没有大段 DNA 丢失或重排的进一步证明来自对有丝分裂期间凝聚的染色体中可观察到的详细带型的比较（图 4-11）。根据这一标准，人体所有分化细胞的染色体组似乎都是一样的。而且，基于重组 DNA 技术对不同细胞基因组的比较表明，作为一般规则，作为多细胞生物发育基础的基因表达并未随着相应基因 DNA 序列的改变而改变。然而，也有一些例子，在生物发育期间发生了基因组 DNA 重排，最著名的就是哺乳动物免疫系统多样性的产生（讨论见第 24 章）。

不同类型的细胞合成不同系列的蛋白质

作为了解细胞分化的第一步，我们愿意去知道任何两个类型细胞之间究竟存在多少差别。尽管我们还不知道这个基本问题的答案，但我们可以做出大致的陈述。

1. 许多过程在所有细胞中都能见到，因而同一生物任何两个细胞拥有许多共同的蛋白质。这些蛋白质包括染色体结构蛋白、RNA 聚合酶、DNA 修复酶、核糖体蛋白，参与新陈代谢主要反应的酶以及许多形成细胞骨架的蛋白质。

2. 有些蛋白质在其执行功能的特殊细胞中含量丰富，却在其他地方检测不到，即使使用灵敏的试验。例如，血红蛋白只能在红细胞中检测到。

3. 对不同细胞中 mRNA 种类的研究表明，在任何时候，一个典型的人体细胞表达全部约 30 000 个基因中的 10 000～20 000 个。如果将一系列不同人体细胞系中的 mRNA 模式进行比较，就会发现在不同细胞类型之间几乎每个活性基因表达水平都不一样。但是大多数差别要细微得多。每种细胞类型都有其特有的 mRNA 丰度模式（使用 DNA 微阵列确定，讨论见第 8 章），所以我们可以用它来对不明组织来源的人类癌细胞进行分类（图 7-3）。

4. 虽然特化细胞类型中 mRNA 的差异很显著，可是它们低估了蛋白质合成模式的全部差别。本章我们将会看到，在转录之后会有很多阶段可以对基因表达进行调控。此外，选择性剪接能从单个基因生成一系列蛋白质。最后，蛋白质合成以后能发生共价修饰。一种更好的显示不同类型细胞基因表达的基本差异的方法是通过二维凝胶电泳，利用这种方法可以直接测量蛋白质水平并显示一些最常见的翻译后修饰（图 7-4）。

细胞能够改变基因的表达以应答外部信号

在多细胞生物中大多数特化细胞都能改变它们的基因表达模式以响应外部信号。例如，如果一个肝细胞接触糖皮质激素后，几种特殊蛋白质的合成量就会显著增加。体内糖皮质激素在饥饿或者剧烈运动时分泌，为肝脏提供信号，增加从氨基酸和其他小分子合成葡萄糖的量；诱导合成的一组蛋白质中包括酶，例如，酪氨酸转移酶，它帮助将酪氨酸转变成葡萄糖。当这种激素不再存在时，这些蛋白质合成量就降到正常水平。

其他细胞类型对糖皮质激素作出不同的响应。例如，在脂肪细胞中，酪氨酸氨基转移酶合成量减少，而一些其他细胞却根本不对糖皮质激素做出响应。这些例子证明了细胞分化的一般特征：不同的细胞类型常常以不同的方式对相同细胞外信号做出响应。响应胞外信号而做出这种调整的基础在于，基因表达模式存在多种特征，它们不发生改变且赋予每种细胞类型永久的与众不同的特点。

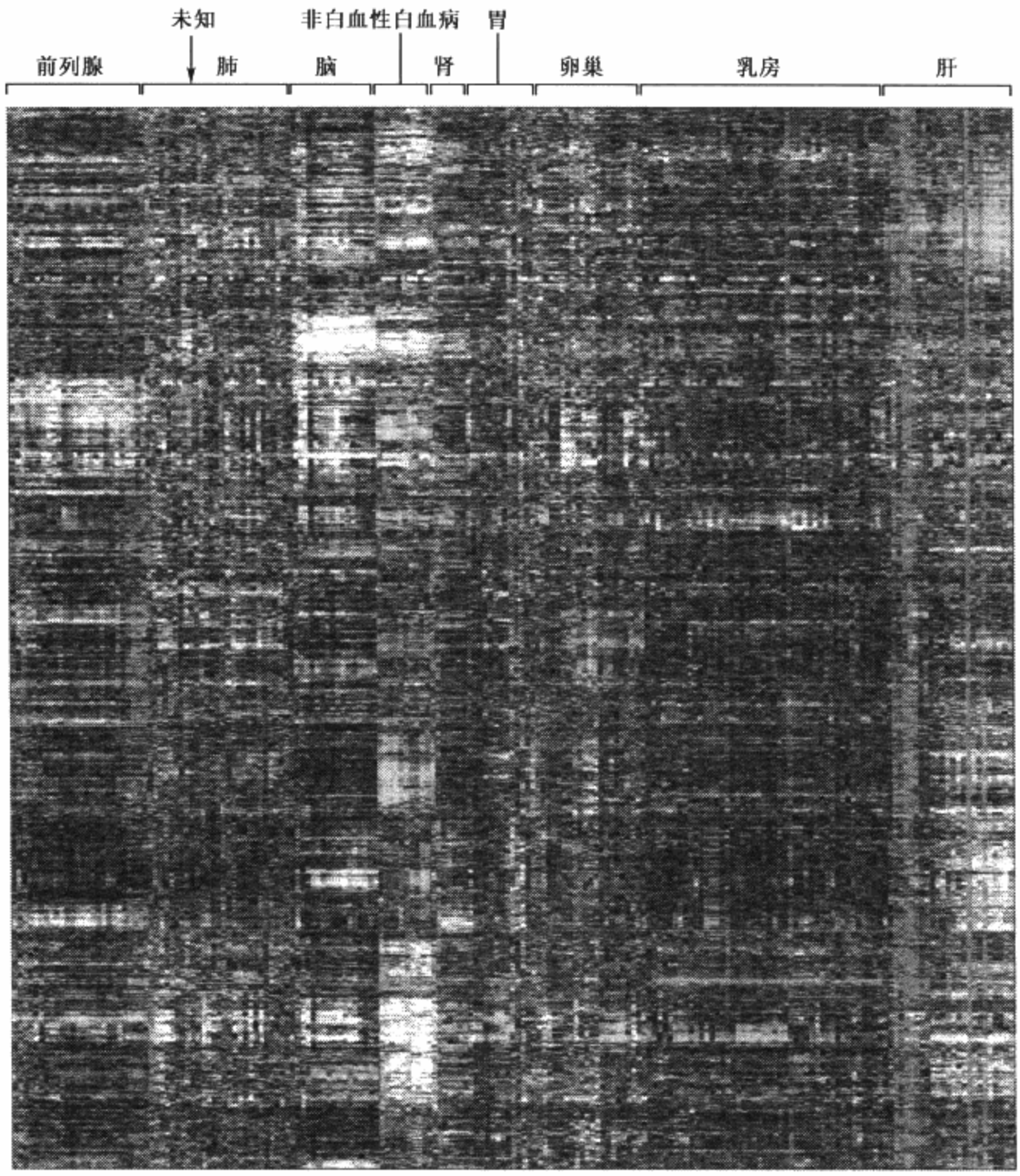


图 7-3　不同的人癌细胞基因表达模式的差别

该图总括了很大一组数据，确定了来自 142 种不同人类肿瘤（从左至右排列）中 1800 个经过选择的基因（从上至下排列）的 mRNA 水平高低，其中每一种肿瘤都来自不同的患者。每个小红条表示给定的基因在给定的肿瘤中转录水平显著高于整个细胞系平均水平；每个小绿条表示低于平均水平。用于生成这些数据的方法（分离 mRNA 后与 DNA 微阵列杂交）在第 8 章中描述。该图表明所分析的 1800 个基因中每一个的相对表达水平在不同的肿瘤中发生变化（可以跟着给定的基因从左向右横跨这幅图看到）。这种分析也显示每一种肿瘤有其独特的基因表达模式。这个信息可以用于确定未知组织来源的癌细胞的类型，通过将基因表达概况同已知肿瘤的表达情况匹配。例如，图中未知的样品经鉴定是肺癌（经 Patrick O. Brown, David Botstein and the Stanford Expression Collaboration 许可）

从 DNA 到 RNA 到蛋白质这条途径中，基因表达可以在许多步骤上受到调控

如果一种生物不同细胞类型之间的差别依赖于细胞表达的特定基因，那么在什么水平上对基因表达进行调控的呢？我们在最后一章将会看到，在从 DNA 到蛋白质的途径中，要经历很多步骤，从理论上讲它们都可以被调节。因此一个细胞能够调控它合成的蛋白质，通过

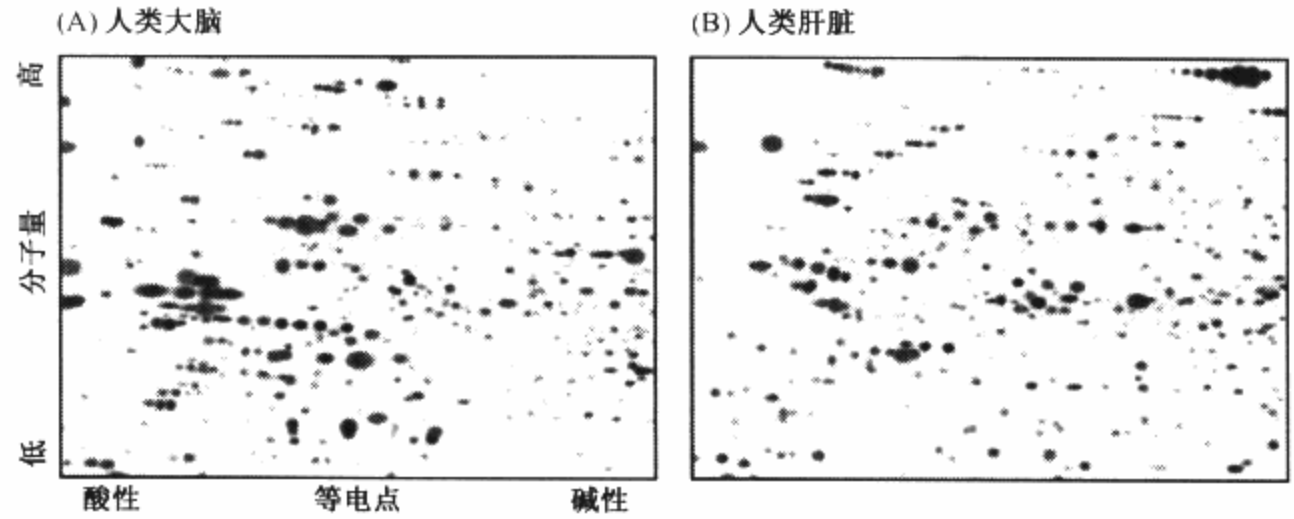


图 7-4　两种人类组织中蛋白质表达的差异

每一块板中，利用二维聚丙烯酰胺凝胶电泳展示这些蛋白质。蛋白质根据相对分子质量（从上至下）和等电点的不同而分开。后者也就是蛋白质不带净电荷时的 pH（从右至左）。人为涂成红色的蛋白质点是两种样品共有的；而蓝色的是其中一种组织所特有的。这两种组织样品之间的差别远比它们的共同点重要：即便对于这两种组织共有的蛋白质而言，它们的相对丰度通常是不同的。注意这种技术利用大小和电荷将蛋白质分开；因而一个蛋白质，如有几种磷酸化状态的蛋白质将呈现出一系列水平斑点（见右板的右上部分）。每个样品只显示完整蛋白质谱的一小部分（经 Tim Myers and Leigh Aderson，Large Scale Biology Corporation 许可）

①调控某个基因转录的时间和频率（**转录调控**）；②调控 RNA 转录物的剪接或其他的加工（**RNA 加工调控**）；③选择将细胞核内哪一条已完成的 mRNA 转移到细胞质并决定将它们定位在细胞质的什么地方（**RNA 转运和定位调控**）；④选择由核糖体对哪条 mRNA 进行翻译（**翻译调控**）；⑤选择性地让细胞质内某些 mRNA 分子失去稳定性（**mRNA 降解调控**）；⑥在特定蛋白质分子合成以后，选择性地活化、失活、降解或区隔它们（**蛋白质活性调控**）（图 7-5）。

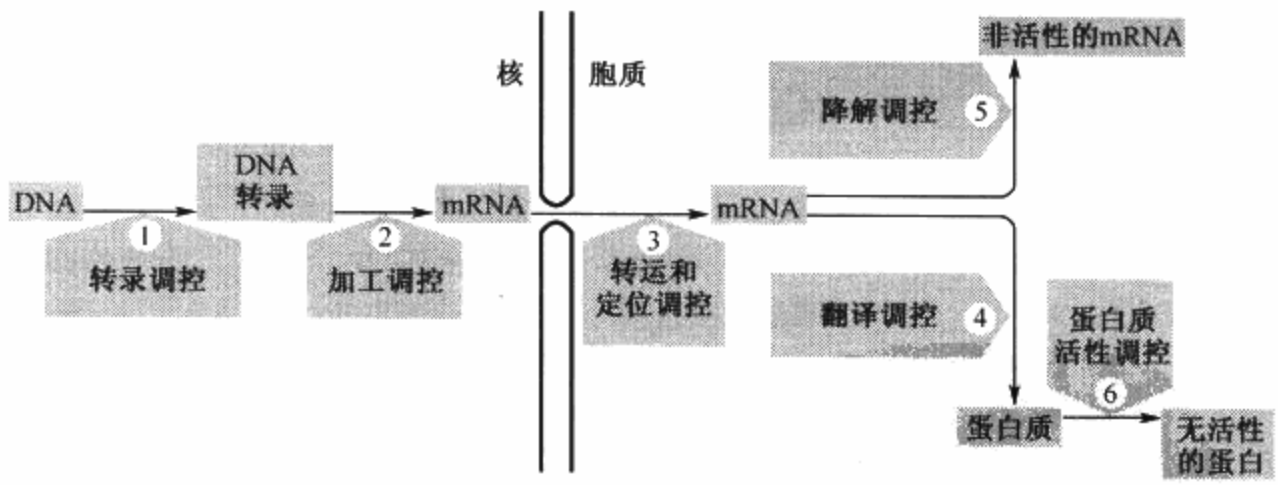


图 7-5　真核基因表达过程中有 6 个步骤可以受到调控

本章讨论了第 1 步到第 5 步进行的调控。第 6 步，蛋白质活性的调节，包括蛋白质磷酸化的可逆激活或者失活（讨论见第 3 章）以及蛋白质水解降解的不可逆失活（讨论见第 6 章）

对于大多数基因，转录调控是极其重要的。这样做是有意义的，因为在图 7-5 所示的所有可能的调控点中，只有转录调控确保不会合成过量的中间产物。在下面的章节中我们讨论通过调节基因转录起始行使这种功能的 DNA 和蛋白质组分。我们将在本章最后再讨论调节基因表达的其他方式。

小结

细胞基因组 DNA 序列中含有合成数千种各种蛋白质和 RNA 分子的信息。一个细胞典型地只表达基因的一部分，多细胞生物中不同类型细胞的出现是因为表达了一组不同的基因。而且，细胞能根据环境的变化，如来自其他细胞的信号，改变它们的基因表达模式。虽然基因表达涉及的所有步骤在理论上都能被调节，但是对于大多数基因来说，RNA 转录的起始是最重要的调节点。

基因调节蛋白中的 DNA 结合基序

细胞是如何决定转录它成千上万个基因中的哪些基因呢？在第 4 章和第 6 章曾简要地提过，各个基因的转录是由离转录起始位点较近的 DNA 调节区域调控的。有些调节区域很简单，就像一个开关，可由单个信号开或关。许多其他的调节区域则很复杂，就像是微处理器，响应不同的信号，对它们进行解释和整合，以便将邻近的基因打开或者关闭。无论复杂还是简单，这些开关装置含有两种基本成分：①具有确定序列的一小段 DNA；②识别并结合这段 DNA 的基因调节蛋白。

我们开始讨论基因调节蛋白时，先来说说它们是如何被发现的。

运用细菌遗传学发现了基因调节蛋白

20 世纪 50 年代在细菌中进行的遗传分析提供了基因调节蛋白存在的第一个证据，这些蛋白质打开或关闭特定的一套基因。其中一个调节蛋白就是 λ 阻遏蛋白，它由一种细菌病毒 λ 噬菌体编码。这个阻遏蛋白将编码新病毒颗粒蛋白组分的病毒基因关闭，因而使病毒基因组像一位安静的乘客一样保留在细菌染色体中，当条件对细菌生长有利时，就随细菌一同增殖（图 5-81）。λ 阻遏蛋白是最先确定特征的基因调节蛋白之一，它也一直是了解最清楚的基因调节蛋白之一，我们将在后面讨论。其他的细菌调节蛋白根据营养条件的变化进行调节，在不需要一组特定的代谢酶时，便将编码它们的基因关闭。作为这些细菌蛋白中最先识别出来的 Lac 阻遏蛋白，在培养基中缺乏乳糖时，将负责乳糖代谢的蛋白质的合成关闭。

了解基因调节的第一步就是要分离出不能关闭特定一套基因的细菌和 λ 噬菌体突变株。当时有人提出，后来也得到了证实，大部分这样的突变体都是缺乏作为那些基因特异阻遏蛋白的蛋白质。与大多数基因调节蛋白一样，这些蛋白质的量很少，所以分离它们很困难，也很费时间。最终通过分级细胞抽提物后纯化得到它们。分离出来以后，显示出这些蛋白质结合在它们所调节的基因邻近的特异 DNA 序列上。然后通过经典遗传学重组试验，DNA 测序和 DNA 足迹试验确定了它们所识别的 DNA 的精确序列（讨论见第 8 章）。

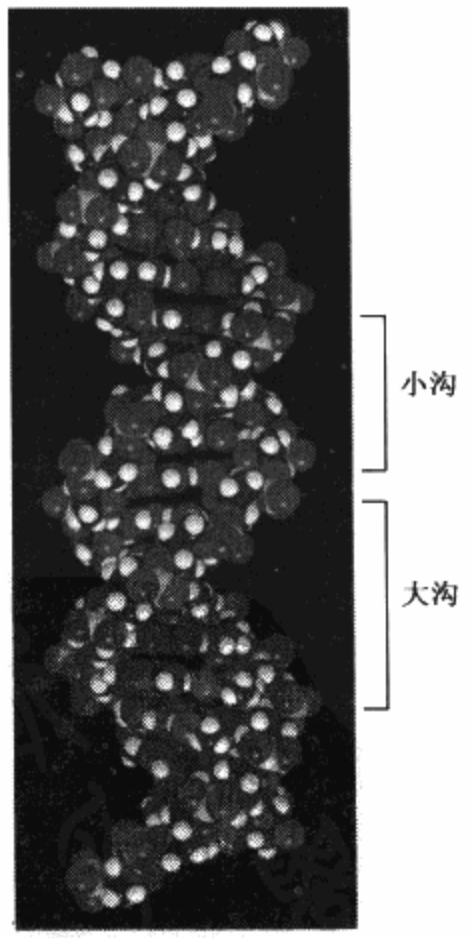


图 7-6　DNA 双螺旋结构

图中显示了位于双螺旋外侧的大沟和小沟。原子分别按下面的说明着色：碳，深蓝色；氮，淡蓝色；氢，白色；氧，红色；磷，黄色

DNA 螺旋的外侧可被蛋白质识别

在第 4 章讨论过，染色体中的 DNA 由一条很长的双螺旋组成（图 7-6）。基因调节蛋白必须识别深埋在这个结构中的特定核苷酸序列。最初人们认为为了区别两条 DNA 序列，可能需要蛋白质直接接触双螺旋内部碱基对之间的氢键。现在已经清楚，双螺旋外部就点缀有 DNA 序列信息，基因调节蛋白可以识别它们而不必打开双螺旋。每个碱基对的边缘都暴露在双螺旋的表面，不论在大沟还是小沟里都呈现出氢键供体、氢键受体，以及疏水区域各具特征的模式，供蛋白质识别（图 7-7）。但是，对于 4 种碱基对的排列来说，只有位于大沟的模式显著不同（图 7-8）。因为如此，基因调节蛋白一般都结合到大沟上，这些我们将会看到。

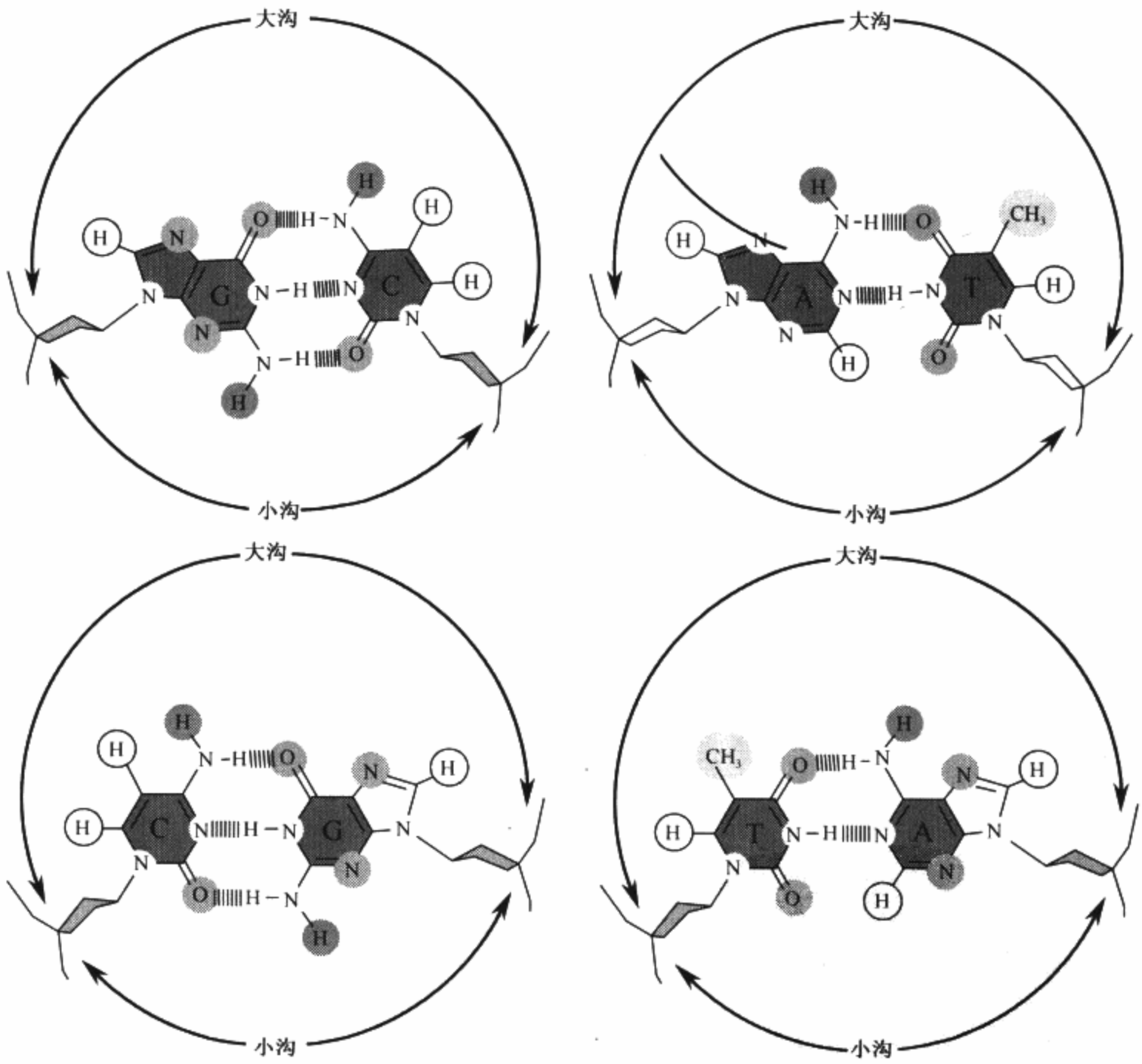


图 7-7　怎样从 DNA 中不同碱基的外侧识别它们而又不必打开双螺旋

图中显示了碱基对的 4 种可能的构型，其中蓝色表示潜在的氢键供体，红色表示潜在的氢键受体，一系列红色的短平行线表示碱基对自身的氢键。形成疏水突出物的甲基用黄色表示，与碳原子连接因而不能用来形成氢键的氢原子用白色表示（摘自 C. Branden and J. Tooze, Introduction to Protein Structure, 2nd edn, New York: Garland Publishing, 1999）

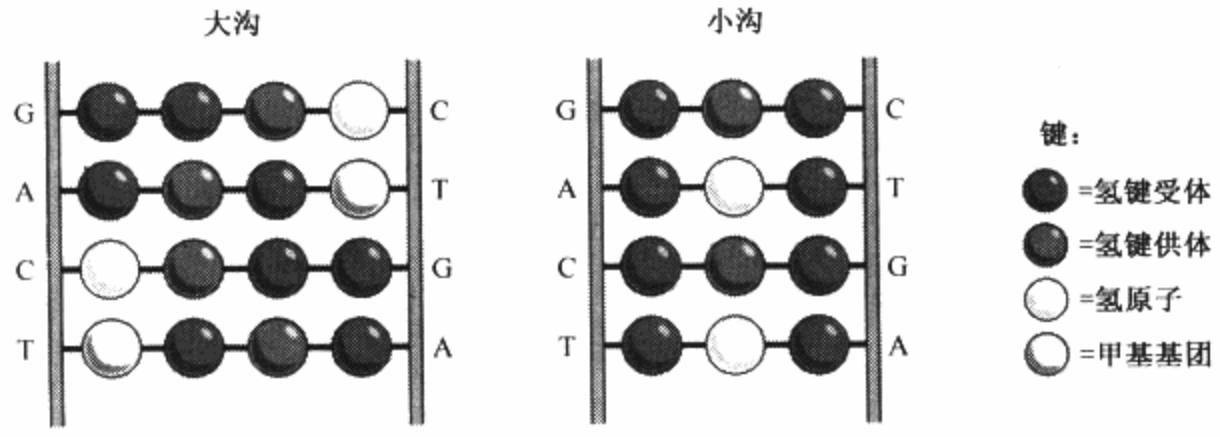


图 7-8　DNA 识别密码

这里直接观察所看到的每个碱基对边缘含有不同的氢键供体、氢键受体和甲基基团模式。从大沟看，4 个碱基对构型每个都投射出独特的模式特征。然而，从小沟看，G-C 和 G-G 的模式相似，A-T 和 T-A 的模式相似。颜色编码与图 7-7 中的相同。（摘自 C. Branden and J. Tooze, Introduction to Protein Structure, 2nd edn, New York: Garland Publishing, 1999）

虽然氢键供体和受体基团模式是基因调节蛋白识别的最重要特征，但它们并不是唯一的：核苷酸序列也决定了双螺旋整体几何学特征，形成了能被识别的“理想化”螺旋扭曲。

DNA 双螺旋的几何学特征由核苷酸序列决定

1953 年发现 DNA 双螺旋之后，20 年来人们一直认为 DNA 是单一的结构，每相邻核苷酸对之间的螺旋扭曲恰好为 36°（每圈螺旋有 10 个核苷酸对）且螺旋几何也一致。然而，这种观点基于对 DNA 分子异源混合物的结构研究而来，一旦使用 X 射线晶体学和核磁共振光谱学分析得到有确定核苷酸序列的 DNA 小分子的三维结构以后，这个观点就改变了。然而早期的研究提供的是一幅平均化、理想化的 DNA 分子图，后来的研究表明，任何给定的核苷酸序列都有局部不规则性。例如，倾斜的核苷酸对或螺旋扭角大于或小于 36°。这些特征也能被特定的 DNA 结合蛋白所识别。

在核苷酸序列中出现的一种特别显著的偏离平均结构的情况，引起 DNA 双螺旋发生弯曲。有些序列（如 AAAANNN，这里 N 可以是除 A 以外的任何碱基）形成一个具有显著不规则性的双螺旋，引起微小的弯曲；如果这个序列在长 DNA 分子中每隔 10 个核苷酸对重复一次，那么这些小弯曲就加合起来，以至于在电子显微镜下观察时可以看到 DNA 分子呈现异常的曲线状（图 7-9）。

DNA 结构的一个相关且同等重要的可变特征就是双螺旋发生变形的程度。一个蛋白质要识别并结合特异的 DNA 序列，该 DNA 和蛋白质之间就必须存在完好的匹配，并且通常必须使正常的 DNA 构象发生扭曲，以便让这种匹配性最大化（图 7-10）。这种扭曲的能量消耗依局部核苷酸序列而定。第 4 章讨论核小体组装时我们遇到过一个这样的例子：有些 DNA 序列能比其他的序列更好地容纳核小体形成时所需要的 DNA 紧密缠绕。同样，许多基因调节蛋白在结合 DNA 时能诱导它发生明显的弯曲（图 7-11）。一般而言，这些蛋白质识别容易弯曲的 DNA 序列。

短 DNA 序列是遗传开关的基本组分

我们已经看到一段特异的核苷酸序列作为 DNA 双螺旋表面结构特征的模式怎样被识别。

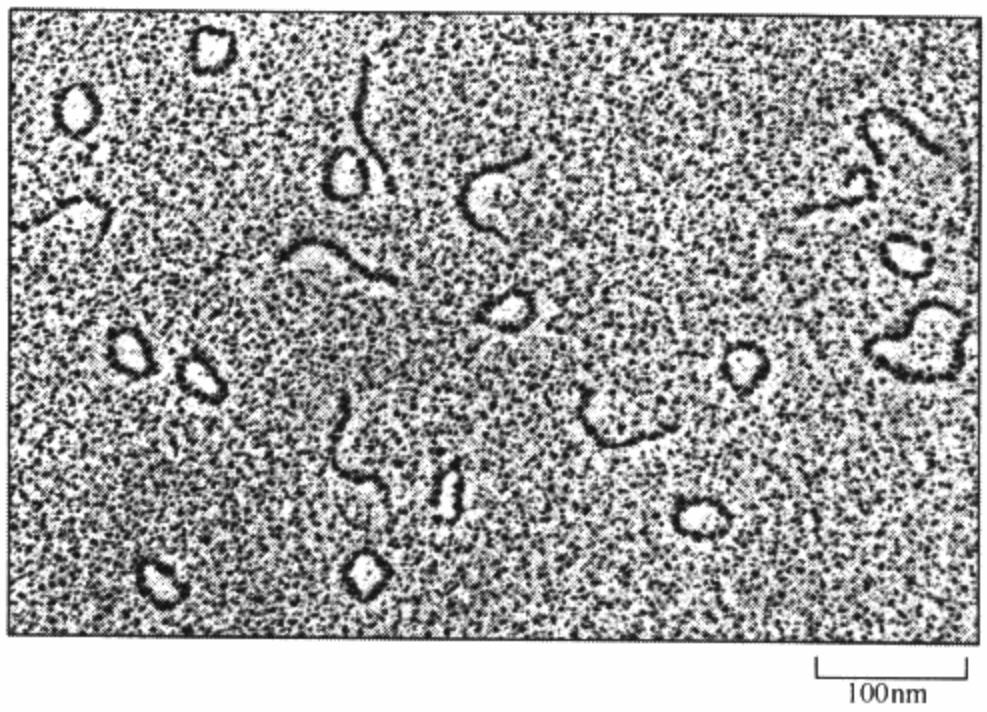


图 7-9　高度曲折的 DNA 双螺旋片段电镜显微照片

这些 DNA 片段来自锥虫的线粒体 DNA 环形小分子。尽管这些片段仅有 20 个核苷酸对，但它们中有许多弯曲后形成完整的圆环。平均来讲，这个长度的正常 DNA 螺旋只能弯曲形成 1/4 的环（一种平滑的右转角）(摘自 J. Griffith, M. Bleyman, C. A. Raugh, P. A. Kitchin, and P. T. Englund, *Cell* 46: 717～724, 1986 © Elsevier)

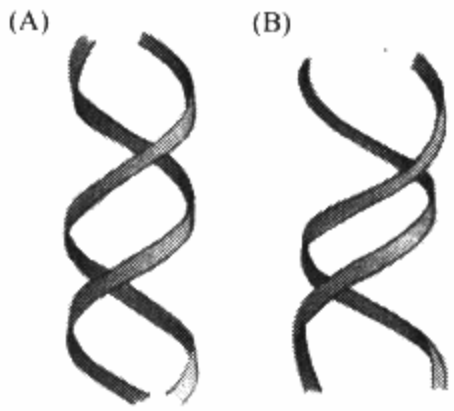


图 7-10　蛋白结合诱导的 DNA 形变

该图显示当一种经过详细研究的基因调节蛋白（噬菌体 434 阻遏蛋白，λ 阻遏蛋白的近亲）结合特异的 DNA 序列时观察到 DNA 结构变化，从常规的双螺旋（A）变为（B）中的扭曲形式。DNA 序列容易发生形变常常影响蛋白质结合的亲和力

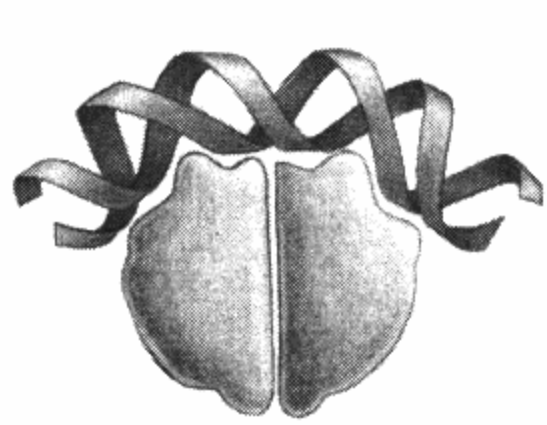


图 7-11　代谢物激活蛋白（CAP）的结合诱导 DNA 弯曲

CAP 是一种来自大肠杆菌的基因调节蛋白，不存在这个结合蛋白时，DNA 螺旋是笔直的

特定的核苷酸序列，通常每个长度不到 20 个核苷酸对，作为特定基因调节蛋白的识别结合位点，其功能相当于遗传开关的基本组分。已经鉴定出来数千个这样的 DNA 序列，每一个都由不同的基因调节蛋白（或者一套相关的基因调节蛋白）识别。表 7-1 中列出了本章所讨论的一部分调节蛋白，以及它们识别的 DNA 序列。

现在我们来看看基因调节蛋白本身，也就是遗传开关第二个基本组分。我们首先讨论使这些蛋白质能够识别包含在一段很长的双螺旋中的短小且特异的 DNA 序列的结构特征。

表 7-1　某些基因调控蛋白调控的 DNA 序列

	名称	识别的 DNA 序列
细菌	lac 阻遏蛋白	5′AATTGTGAGCGATAACAATT
		3′TTAACACTCGCCTATTGTTAA
	CAP	TGTGAGTTAGCTCACT
		ACACTCAATCGAGTGA
	lambda 阻遏蛋白	TATCACCGCCAGAGGTA
		ATAGTGGCGGTCTCCAT
酵母	Gal4	CGGAGGACTGTCCTCCG
		GCCTCCTGACAGGAGGC
	Matα2	CATGTAATT
		GTACATTAA
	Gcn4	ATGACTCAT
		TACTGAGTA
果蝇	Kruppel	AACGGGTTAA
		TTGCCCAATT
	Bicoid	GGGATTAGA
		CCCTAATCT
哺乳动物	Sp1	GGGCGG
		CCCGCC
	Oct-1 Pou 结构域	ATGCAAAT
		TACGTTTA
	GATA-1	TGATAG
		ACTATC
	MyoD	CAAATG
		GTTTAC
	p53	GGGCAAGTCT
		CCCGTTCAGA

注：该表中的每个蛋白能识别一组非常相似的 DNA 序列（见图 6-12）；为了方便起见，除了保守序列外，每个蛋白只给出一个识别序列。

基因调节蛋白含有能够阅读 DNA 序列的结构基序

生物学中分子识别一般有赖于两个分子表面的恰好适配，基因调节蛋白的研究为这种原理提供了一些最清楚的例子。一个基因调节蛋白识别一段特异的 DNA 序列是因为蛋白质表面与双螺旋上结合区域特殊的表面特征广泛互补。在大多数情况下，蛋白质与 DNA 大量接触，涉及氢键、离子键，以及疏水相互作用。虽然单个接触很微弱，可是通常在蛋白质-DNA 表面形成的 20 个左右的接触加起来确保了这种相互作用既特异又很牢固（图 7-12）。实际上，DNA-蛋白质相互作用包括了一部分生物学中所知道的最牢固和最特异的分子相互作用。

虽然每个蛋白质-DNA 识别的例子在细节上都很独特，对几百种基因调节蛋白进行 X 射线晶体学和 NMR 光谱学研究后，表明许多蛋白质含有一组 DNA 结合结构基序中的某一小部分。这些基序一般利用 α 螺旋或 β 折叠结合到 DNA 的大沟上；我们已经看到了，这个大沟含

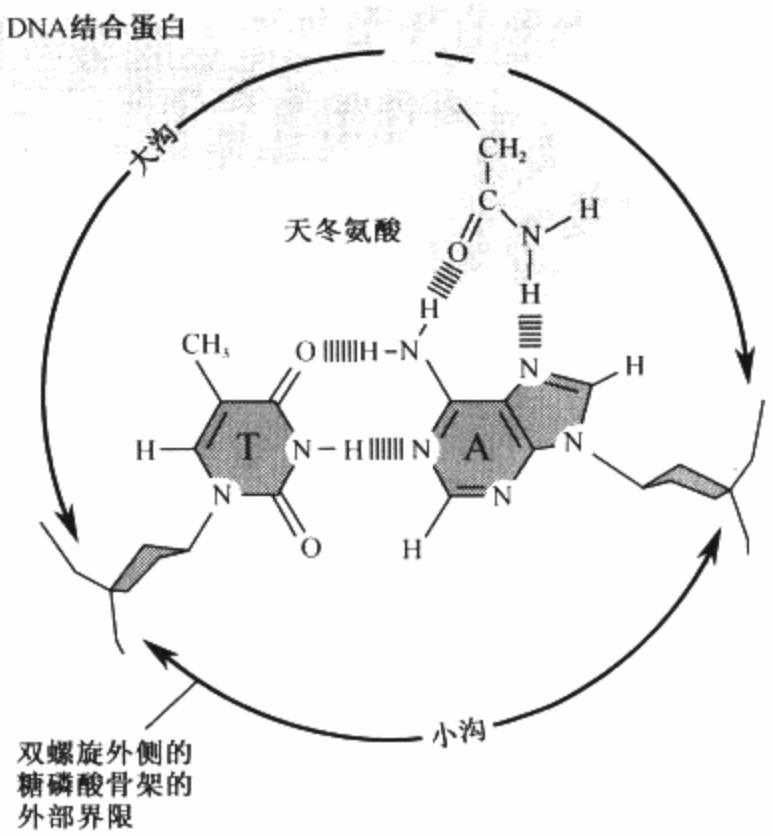


图 7-12　基因调节蛋白对 DNA 大沟的结合

图中只显示了一种接触，DNA-蛋白质交界面典型地由 10～20 种这样的接触组成，其中涉及不同的氨基酸，每一种都为蛋白质-DNA 相互作用做出贡献

有足以区别不同 DNA 序列的信息。这种适配是如此完美，以至于有人提出核酸基本结构单位和蛋白质的大小一起进化，从而允许这些分子能相互连锁起来。

螺旋-转角-螺旋是最简单也是最常见的 DNA 结合基序之一

第一个需要认识的 DNA 结合蛋白基序是螺旋-转角-螺旋。这种基序最初在细菌蛋白中鉴定出来，自那以后在来自原核和真核生物的数百种 DNA 结合蛋白质中都有发现。它由一条形成“转角”的伸展氨基酸短链连接两个 α 螺旋构成（图 7-13）。这两个螺旋主要通过两个螺旋之间的相互作用以固定的角度维系起来。更接近 C 端的螺旋称为识别螺旋，因为它结合到 DNA 大沟中；它的氨基酸侧链因不同的蛋白质而存在差别，在识别蛋白结合到特异 DNA 序列时发挥重要作用。

在螺旋-转角-螺旋区域的外侧，含有这种基序的不同蛋白质的结构变化很大（图 7-14）。因此每种蛋白质以独特的方式将它的螺旋-转角-螺旋基序呈现给 DNA，这个特征被认为通过增加了利用该基序识别的 DNA 序列数量，来增加螺旋-转角-螺旋基序的功能多样性。此外，在大多数这样的蛋白质中，位于螺旋-转角-螺旋区域外部的多肽链部分也与 DNA 产生重要的接触，帮助调整这种相互作用。

图 7-14 示出的一组螺旋-转角-螺旋蛋白质说明了许多序列特异的 DNA 结合蛋白共有的一种特征。它们作为对称的二聚体结合到同样由两个很相似的对称排列的“半位点”组成的 DNA 序列上（图 7-15）。这种排列使每个蛋白质单体能形成一组几乎完全相同的接触位点，因而极大地增加了结合亲和力：初接近时，接触位点数的加倍会使相互作用自由能加倍，因而使亲和常数呈平方增加。

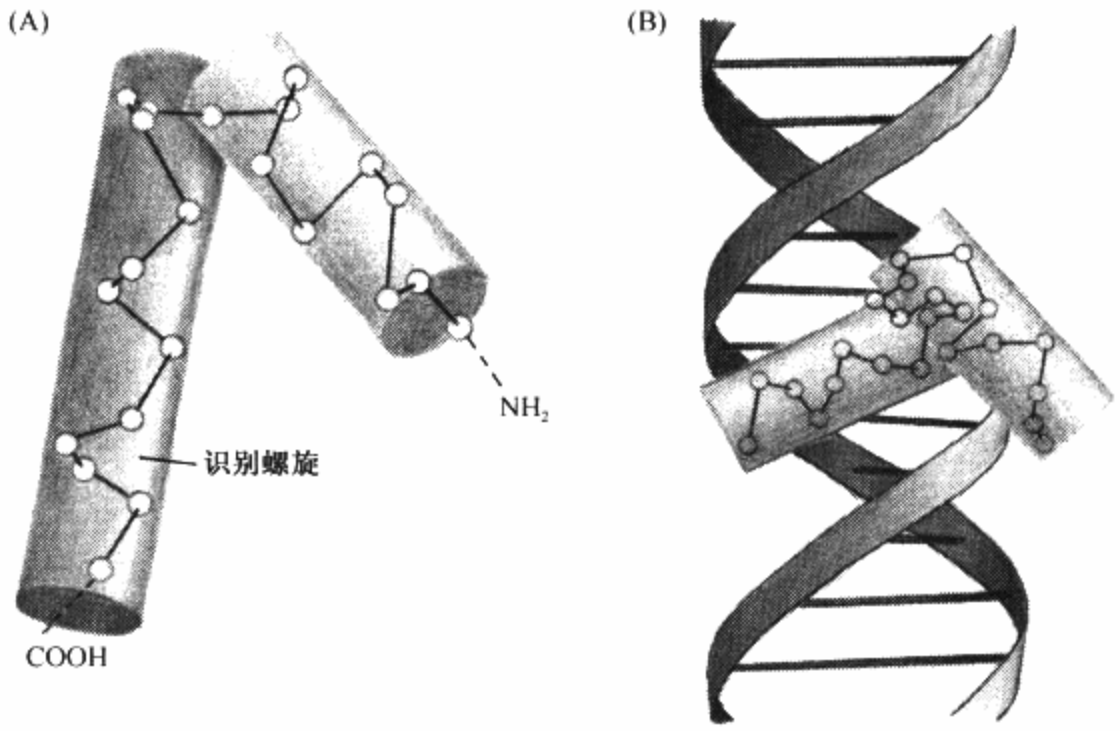


图 7-13　结合 DNA 的螺旋-转角-螺旋基序

（A）中显示了这种基序，其中每个白色圆圈代表氨基酸的中心碳原子。C 端 α 螺旋（红色）称为识别螺旋是因为它参与 DNA 序列特异性识别。如（B）中所示，该螺旋正好结合到 DNA 大沟中，在这里它与碱基对的边缘接触（也见图 7-7）。N 端 α 螺旋（蓝色）的作用主要是作为结构组分帮助定位识别螺旋

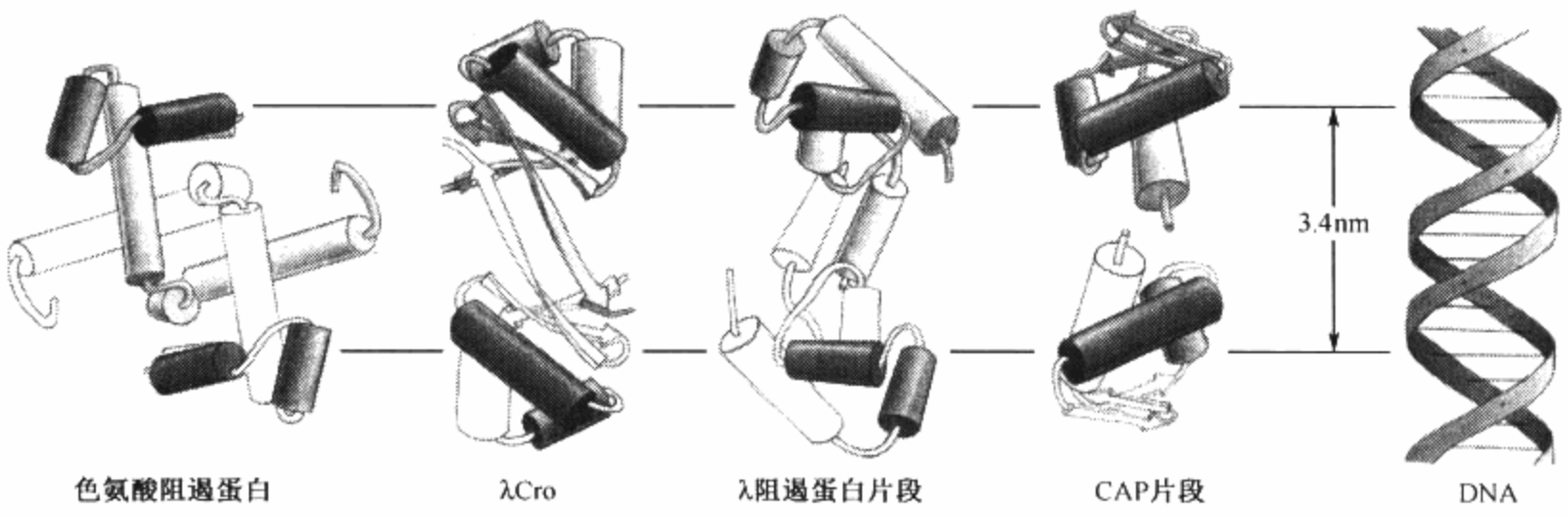


图 7-14　一些螺旋-转角-螺旋 DNA 结合蛋白

所有这些蛋白质以二聚体结合 DNA，在此二聚体中识别螺旋的两份拷贝（红色圆柱）恰好由一圈 DNA 螺旋（3.4nm）分隔开。蓝色表示螺旋-转角-螺旋基序中的另一螺旋，如图 7-13 所示。λ 阻遏蛋白和 Cro 蛋白调控 λ 噬菌体基因表达。色氨酸阻遏蛋白和分解代谢物激活蛋白（CAP）调控成套的大肠杆菌基因表达

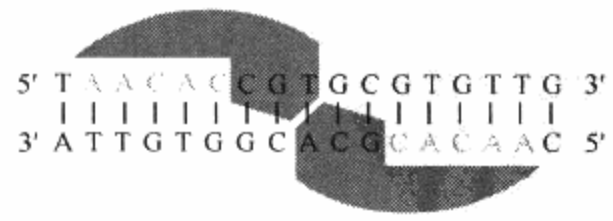


图 7-15　λ 噬菌体 Cro 蛋白识别的特异 DNA 序列

该序列中用绿色标记的核苷酸对称排列，使得 DNA 位点每一半能以相同的方式被每个也用绿色显示的蛋白质单体所识别。该蛋白质的真实结构见图 7-14

同源异型结构域蛋白构成一类特殊的螺旋-转角-螺旋蛋白质

在细菌中发现第一个基因调节蛋白后不久，在果蝇中进行遗传分析导致鉴定出了一类重要基因的特征，它们叫同源异型选择者基因，在使果蝇属发育同步化过程中发挥关键作用。第 21 章将会讨论到，自那以后已经证明它们在高等动物发育中也起根本性的作用。这些基因发生变异导致果蝇身体一部分转变成另一部分，表明它们编码的蛋白质控制了关键的发育决定。

20 世纪 80 年代当鉴定出多个同源异型选择者基因的核苷酸序列后，证明它们每一个都含有一段几乎一模一样的 60 个氨基酸的片段，它们定义了这类蛋白质，并被称为同源异型结构域。当确定了同源异型结构域的三维结构以后，可以看到它含有与细菌基因调节蛋白相关的螺旋-转角-螺旋基序，第一次表明了在细菌中确立的基因调控规则也与更高级的生物有关联。单是在果蝇属中就已经鉴定出 60 多种同源异型结构域蛋白，实际上，从酵母到植物再到人类，所有研究过的真核生物中都已鉴定出了同源异型结构域蛋白。

图 7-16 显示了一个与其特异 DNA 序列结合的同源异型结构域的结构。尽管细菌基因调节蛋白的螺旋-转角-螺旋基序常常包埋在周围不同结构之中，同源异型结构域的螺旋-转角-螺旋基序却通常由相同的结构（它们形成其他的同源异型结构域）环绕，表明这种基序总是以相同的方式呈现给 DNA。事实上，结构研究显示，一种酵母同源异型结构域蛋白与一种果蝇属同源异型结构域蛋白具有非常相似的构象，并以几乎一模一样的方式识别 DNA，尽管它们 60 个氨基酸位点中只有 17 个相同（图 3-15）。

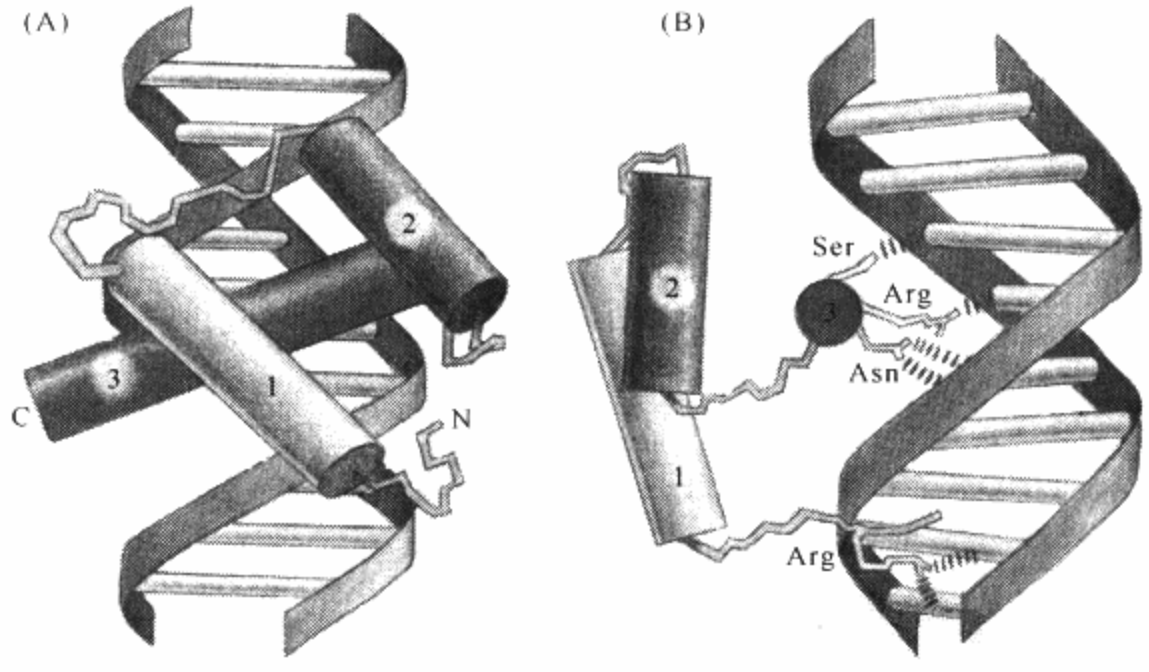


图 7-16　与其特异 DNA 序列结合的同源异型域

图中显示了同一结构的两幅不同的图片。（A）同源异型域折叠成 3 个 α 螺旋，它们通过疏水相互作用紧密地包装在一起。含有螺旋 2 和螺旋 3 的部分有非常相似的螺旋-转角-螺旋基序。（B）识别螺旋（螺旋 3，红色）与 DNA 大沟产生重要的接触。螺旋 3 的天冬酰胺，如与一个腺嘌呤接触，如图 7-12 所示。小沟中的核苷酸对也与螺旋 1 上连接的柔性臂接触。这里所示的同源异型域来自一种酵母基因调节蛋白，但它与许多真核生物的同源异型域都很相似

（摘自 C. Wolberger et al., *Cell* 67：517～528，1991）

存在许多种 DNA 结合锌指基序

螺旋-转角-螺旋基序仅仅由氨基酸组成。另一组重要的 DNA 结合基序增加了一个或多

个锌原子作为结构组分。虽然所有这种与锌协同的 DNA 结合基序都称为锌指，但这种描述仅指最初发现它们以来在简图中的外形[图7-17(A)]。随后的结构研究表明，它们可分为多种不同的结构类型，这里考察其中的两种。第一种类型最初是在激活一种真核生物核糖体 RNA 基因转录的蛋白质中发现。它的结构简单，由锌原子将一个 α 螺旋和 β 折叠维系起来[图 7-17(B)]。这种类型的锌指常常发现与其他锌指位于一个簇中，一个接一个地排列起来，这样每个 α 螺旋就能与 DNA 大沟接触，沿大沟形成一段几乎连续的 α 螺旋。通过这种方式，重复的基本结构单位形成了强有力而且特异的 DNA-蛋白质相互作用（图 7-18）。这种基序的一个特别的优势在于，DNA-蛋白质相互作用的强度和特异性在进化期间可由锌指重复数目的改变进行调整。相比之下，很难想像本章讨论的任何其他 DNA 结合基序怎样能够形成重复的链。

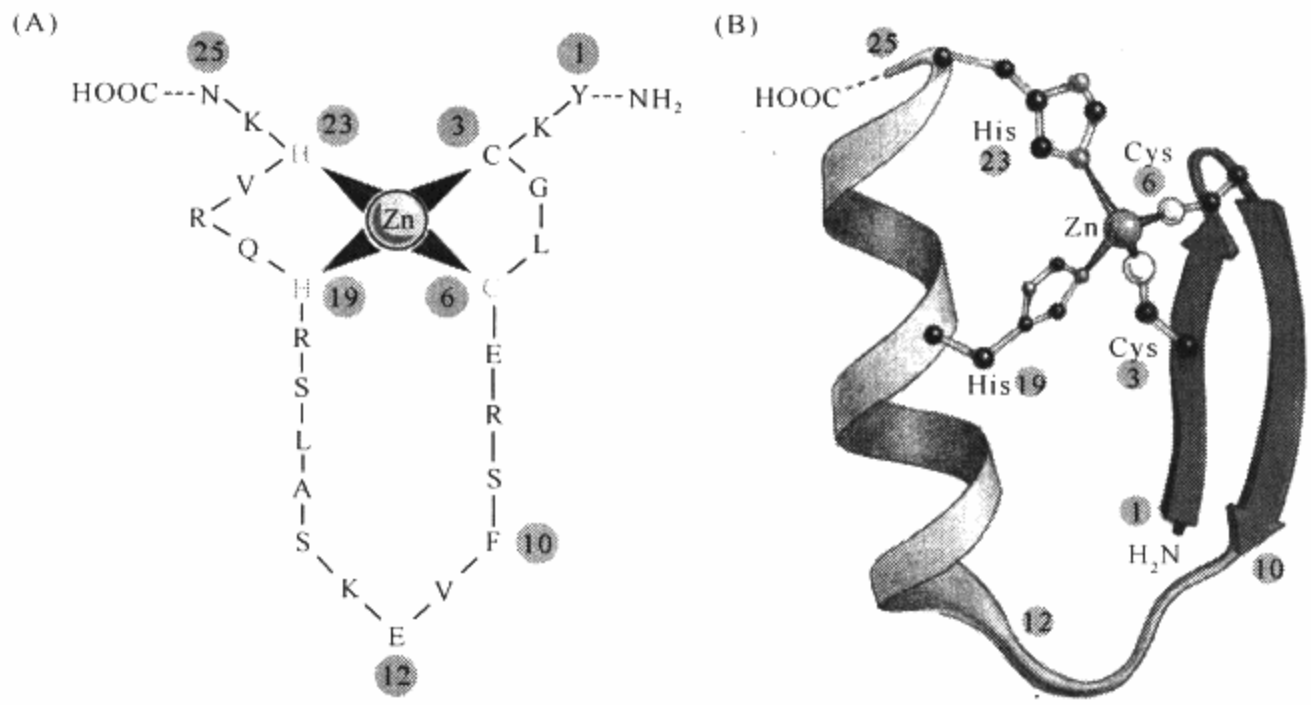


图 7-17　一种锌指蛋白

该蛋白质属于锌指蛋白的半胱氨酸-半胱氨酸-组氨酸-组氨酸家族，是以结合锌的氨基酸来命名。(A) 来自青蛙的这类蛋白质中一个锌指的氨基酸序列简图。(B) 这类锌指的三维结构由反平行的 β 折叠（氨基酸 1 到 10）和紧跟着的 α 螺旋（氨基酸 12 到 24）构成。结合锌的 4 个氨基酸（半胱氨酸 3、半胱氨酸 6、组氨酸 19 和组氨酸 23）将 α 螺旋一个末端牢牢地固定到 β 折叠的一个末端上

（摘自 M. S. Lee et al., *Science* 245：635～637，1989）

另一类锌指在一个细胞内受体蛋白大家族中发现（讨论见第 15 章）。它形成了一种不同类型的结构（在有些方面与螺旋-转角-螺旋基序类似），其中两个 α 螺旋与锌原子一起包装起来（图 7-19）。与螺旋-转角-螺旋蛋白类似，这些蛋白质通常形成二聚体，使每个亚基的两个 α 螺旋中有一个能与 DNA 大沟相互作用（图 7-14）。虽然本部分讨论的两类锌指在结构上不同，但它们拥有两个共同的重要特征：都使用锌原子作为结构元件，都利用 α 螺旋识别 DNA 的大沟。

β 折叠也能识别 DNA

到目前为止讨论的 DNA 结合基序中，α 螺旋是主要的用于识别特异 DNA 序列的机制。然而，有一组基因调节蛋白进化出一种完全不同但同样精巧的识别策略。在这种情况下位于大沟表面的信息由一个两股 β 折叠阅读，该折叠的氨基酸侧链从折叠伸向 DNA，如图 7-20

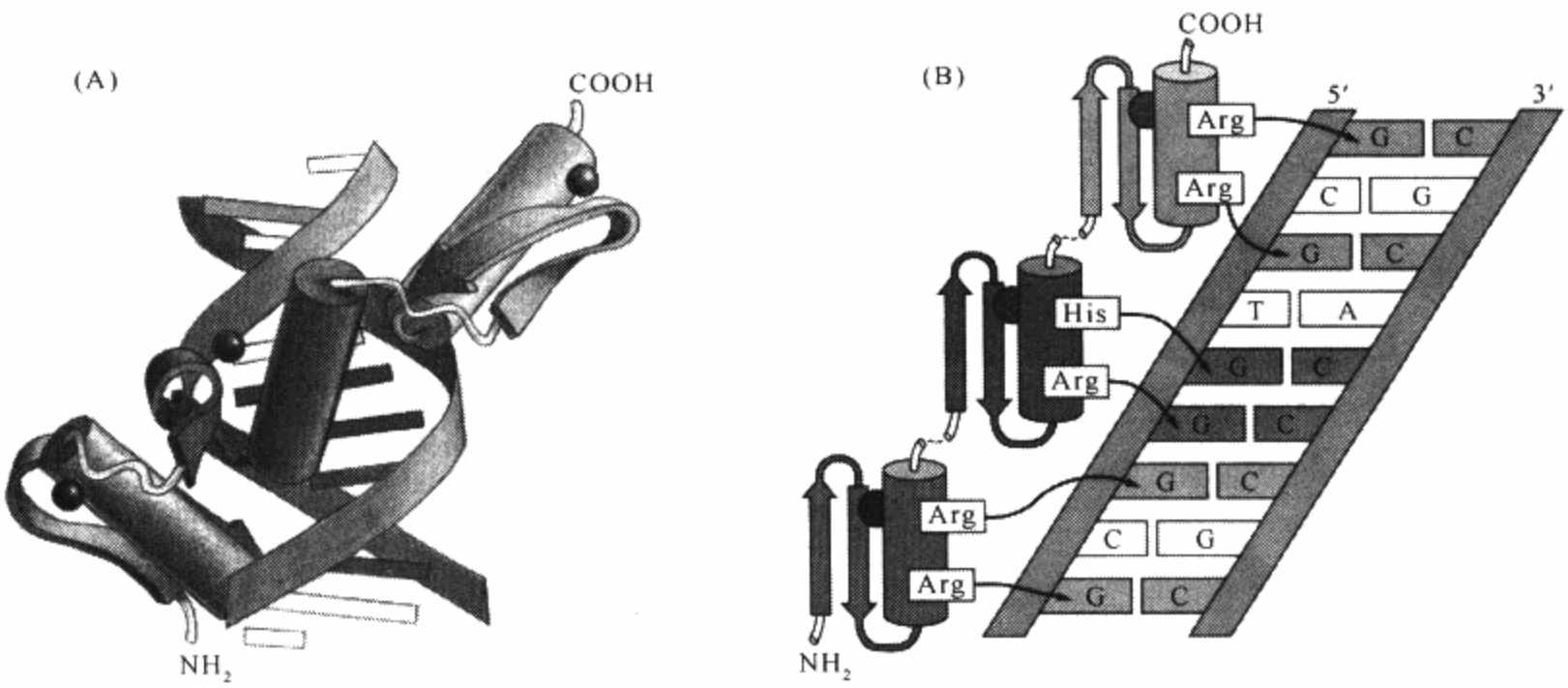


图 7-18　一种锌指蛋白对 DNA 的结合

（A）结合到特异 DNA 位点上的一种小鼠基因调节蛋白片段的结构，该蛋白质利用 3 个直接重复排列的 Cys-Cys-His-His 锌指类型识别 DNA（图 7-17）。（B）这 3 个锌指用于相似的氨基酸序列并以相似的方式接触 DNA。（A）和（B）中每个小球都代表锌指中的锌原子（摘自 N. Pavletich and C. Pabo, *Science* 252：810～817，1991）

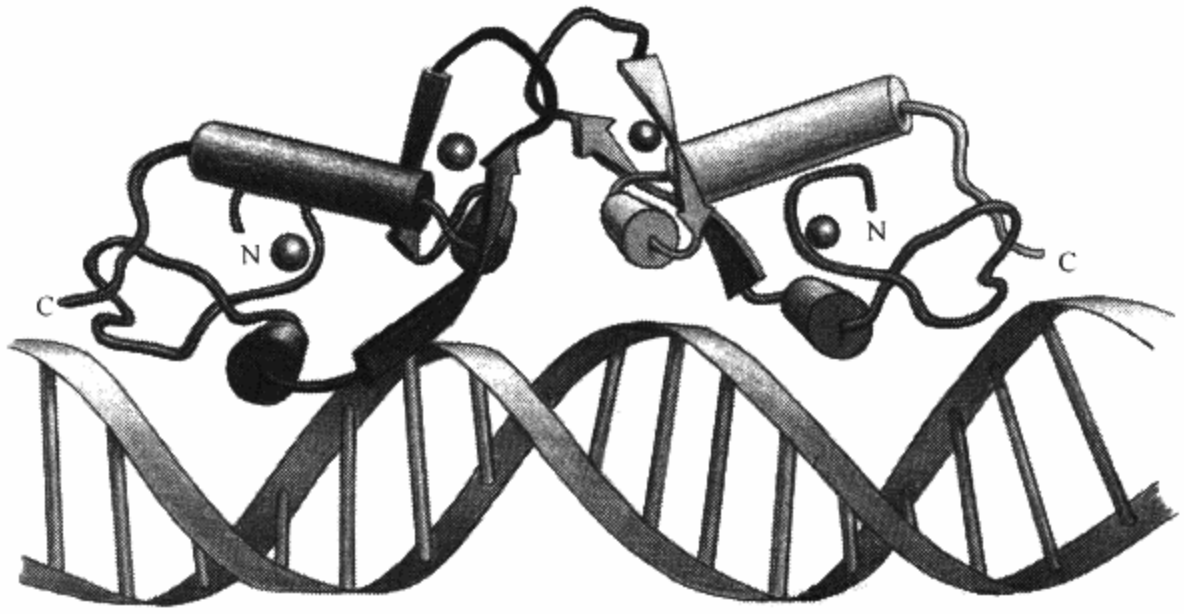


图 7-19　结合其特异 DNA 序列的细胞内受体家族的二聚体锌指结构域

每个锌指结构域含有两个锌原子（用灰色小球表示）：一个稳定 DNA 识别螺旋（其中一个亚单位用褐色表示，另一个用红色表示），另一个稳定一个参与形成二聚体的环（紫色表示）。每个 Zn 原子与 4 个具有适当空间分布的半胱氨酸残基协同作用。与图 7-14 中显示的螺旋-转角-螺旋蛋白相似，二聚体的两个识别螺旋对应于每一圈 DNA 双螺旋分开一定的距离。图中所示的特定例子是糖皮质激素受体片段。这是细胞借以发现并在转录上应答肾上腺应激产生的糖皮质激素的蛋白质（摘自 B. F. Luisi et al., *Nature* 352：497～505，1991）

所示。与 α 识别螺旋情况一样，这种 β 折叠基序也能用于识别许多不同 DNA 序列；确切 DNA 序列的识别由组成 β 折叠的氨基酸序列决定。

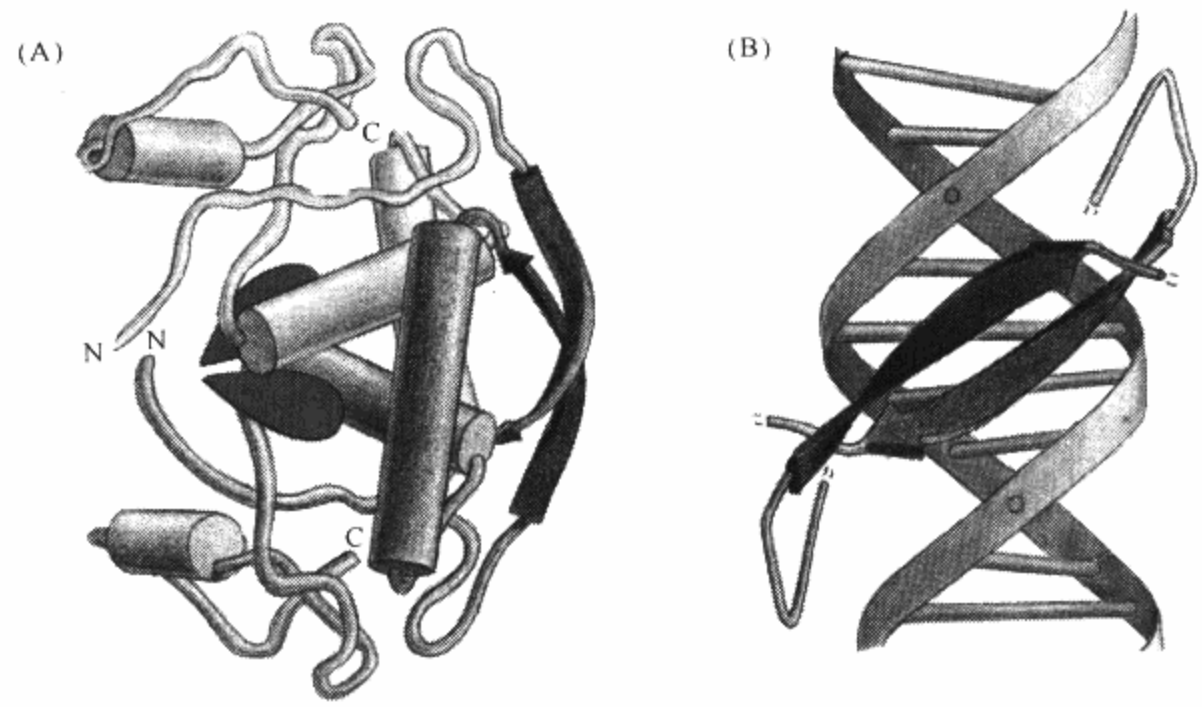


图 7-20　细菌甲硫氨酸阻遏蛋白

细菌甲硫氨酸阻遏蛋白调节催化甲硫氨酸合成的酶的编码基因。当这个氨基酸富余时，它就结合阻遏蛋白，引起蛋白质结构发生变化，从而使它能够牢固结合 DNA，关闭该酶的合成。（A）为了能牢固结合 DNA，甲硫氨酸阻遏蛋白必须与 S-腺苷甲硫氨酸（红色所示）形成复合物。二聚体蛋白的一个甲基用绿色显示，另一个用蓝色显示；结合 DNA 的双链 β 折叠由各来自每个亚基的一条链组成，并分别用深蓝色和深绿色显示。（B）结合到 DNA 上的甲硫氨酸阻遏蛋白的简图，显示了阻遏蛋白的双链 β 折叠如何结合 DNA 的大沟。为了清楚起见，阻遏蛋白其他区域已略去［(A)摘自 S. Phillips. *Curr. Opin. Struc. Biol.* 1：89～98，1991；(B) 摘自 W. Somers and S. Pillips，*Nature* 359：387～393，1992］

亮氨酸拉链基序不仅介导 DNA 结合，而且介导蛋白质二聚化

许多基因调节蛋白以同源二聚体识别 DNA，正如我们已经看到的，可能是因为这是一种获得强有力特异性结合的简单方式（图 7-15）。通常，负责二聚化的蛋白质部分与负责与 DNA 结合的蛋白质部分是不同的（图 7-14）。然而，有一种基序将这两种功能以巧妙而且经济的方式结合起来。它称为**亮氨酸拉链**，之所以这样称谓，是因为分别来自每个单体的 α 螺旋结合起来形成一段短小的卷曲螺旋的方式（图 3-11）。从每个螺旋的一侧伸出的疏水氨基酸侧链（通常在亮氨酸上）之间的相互作用将两个螺旋维系起来，就在两个 α 螺旋二聚化接触面以外，它们彼此分离形成一个 Y 形结构，这使它们的侧链能够接触 DNA 的大沟。因此这个二聚体将双螺旋紧紧扣住，就像晾衣绳上的夹子一样（图 7-21）。

异源二聚化扩大了基因调节蛋白所能识别的 DNA 序列种类

到目前为止我们已看到的许多基因调节蛋白结合 DNA 都是以同源二聚体的形式，即由两个相同的亚基组成的二聚体。然而许多基因调节蛋白，包括亮氨酸拉链，也能与不相同的成分结合起来，形成由两个不同亚基组成的异源二聚体。由于典型的异源二聚体由两种具有不同的 DNA 结合特异性的蛋白质形成，所以基因调节蛋白的混合和配对形成的异源二聚体大大增加了这些蛋白质能够展示的 DNA 结合特异性。如图 7-22 所示，从理论上讲，由两类亮氨酸拉链单体可以形成 3 种不同的 DNA 结合特异性，而 3 类单体则可以形成

图 7-21　结合到 DNA 上的亮氨酸拉链二聚体

两个 α 螺旋 DNA 结合结构域（底部）通过它们的 α 螺旋亮氨酸拉链区域（顶部）二聚化形成一个倒 Y 形结构。Y 的每一条臂由一个 α 螺旋形成，各来自一个单体，并由该臂介导对 DNA 大沟中特定 DNA 序列的结合。每个 α 螺旋结合对称 DNA 结构的一半。图中所示结构是酵母 Gcn4 蛋白，它响应环境中可资利用的氨基酸对转录进行调节（摘自 T. E. Ellenberger et al., *Cell* 71：1223～1237，1992）

6 种，依此类推。

然而这种任意配对也是有限制的：如果一个典型的真核细胞中许多类型的亮氨酸拉链蛋白全部形成异源二聚体，那么细胞的基因调控环路之间的“交叉对话”数量就会如此巨大，以至于会产生混乱。是否能形成一个特定的异源二聚体决定于两个亮氨酸拉链 α 螺旋的疏水表面彼此接合得有多好，继而又决定于两个拉链区域确切的氨基酸序列。因此细胞内每个亮氨酸拉链蛋白只能与其他亮氨酸拉链蛋白中的小部分形成二聚体。

异源二聚体化是组合式调控的一个例子。在这种调控中，由不同的蛋白质组合起来而不是单个蛋白质来调控细胞过程。异源二聚化是真核生物细胞用来调控基因表达的机制之一，它在广泛的不同类型基因调节蛋白中发生（图 7-23）。我们后面还会讨论到，异源二聚化基因调节复合物的形成只是调控基因表达的多种组合机制中的一种。

在基因调节蛋白进化的过程中，类似的组合规则将两个不同的 DNA 结合结构域连接成单个多肽链，从而产生了新的 DNA 结合特异性（图 7-24）。

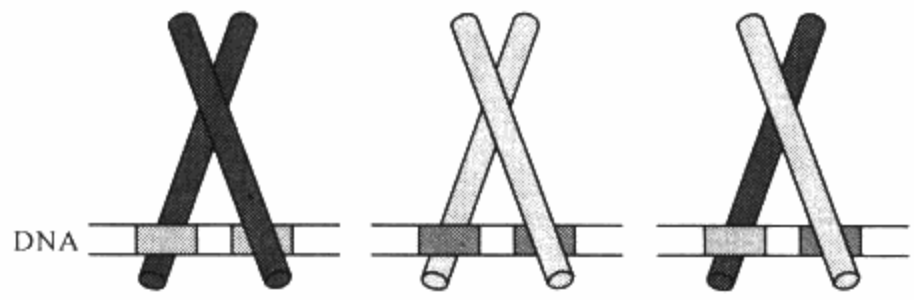


图 7-22　亮氨酸拉链蛋白异源二聚化能改变它们的 DNA 结合特异性

亮氨酸拉链同源二聚体结合对称的 DNA 序列，如左边和中间的图中所示。这两种蛋白质识别不同的 DNA 序列，如图中 DNA 中用红色和蓝色区域所示。两种不同的单体可以组合形成异源二聚体，识别由红色和蓝色区域组成的杂合 DNA 序列

螺旋-环-螺旋也介导二聚化和 DNA 结合

另一种与亮氨酸拉链相关的重要的 DNA 结合基序是螺旋-环-螺旋（HLH），不应当将它和前面讨论的螺旋-转角-螺旋混淆起来。一个 HLH 基序由一个环将一个短 α 螺旋和另一个更长的 α 螺旋连接起来组成。这个环的可变性允许一个螺旋回折跟另一个 HLH 蛋白叠在一起。正如图 7-25 中所示，这个双螺旋结构既可以与 DNA 结合，也可以与另一个 HLH 蛋白

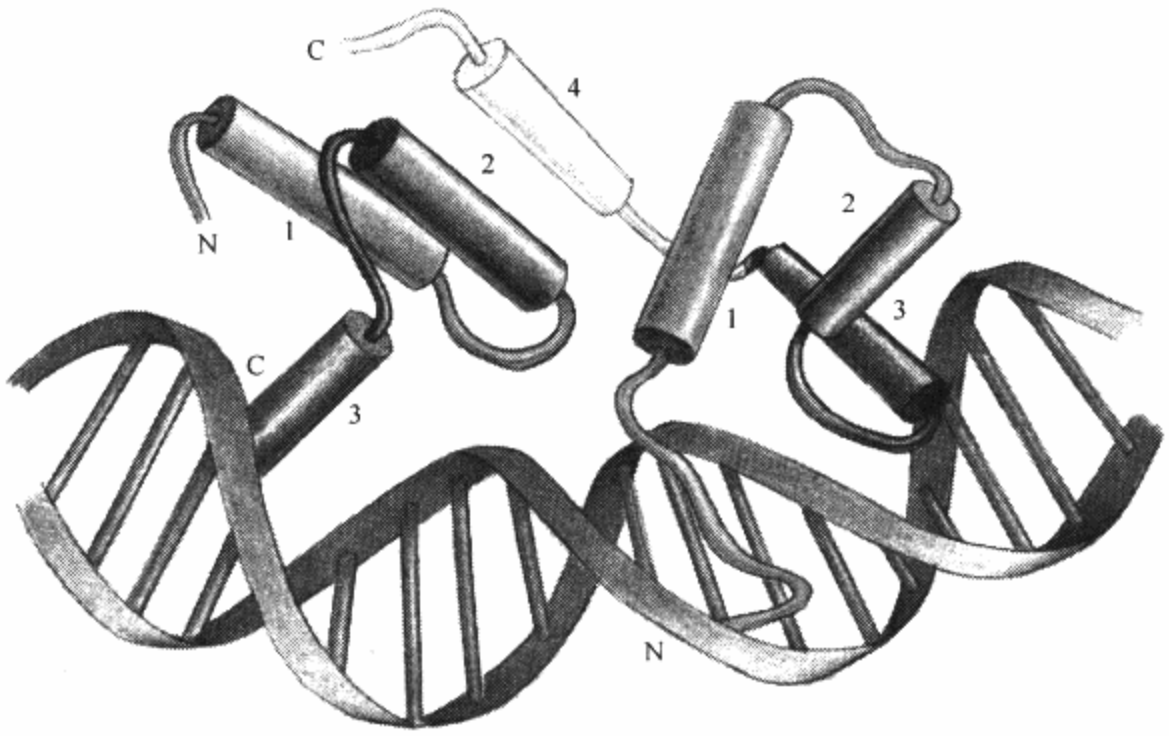


图 7-23　与其 DNA 识别位点结合的两个同源异型域蛋白组成的异源二聚体

右边蛋白质（Matα2）的 4 号黄色螺旋在不存在左边的蛋白质（Matα1）时发生结构破坏，只有在异源二聚化时才形成一个螺旋。因而 DNA 序列由两个蛋白质共同识别；图 7-16 显示了 α2 产生的部分蛋白质-DNA 接触。这两种蛋白质来自出芽酵母，在其中异源二聚体指令了特定的一种细胞类型（图 7-65）。螺旋的编号与图 7-16 中的一致（摘自 T. Li et al., *Science* 270：262～269，1995）

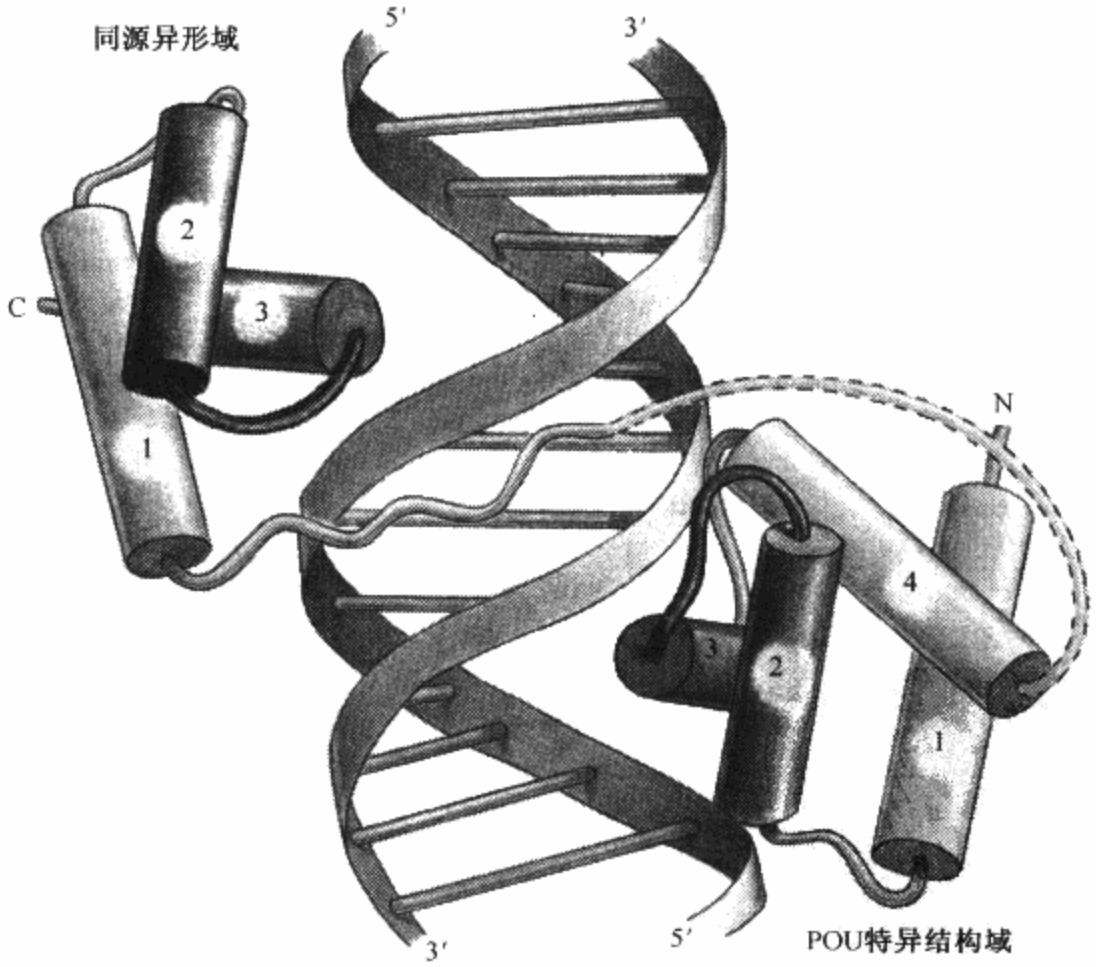


图 7-24　两个 DNA 结合结构域由一条柔性的多肽链共价相连

图中所示的结构（称为 POU 结构域）由一个同源结构域和一个螺旋-转角-螺旋结构（与 λ 噬菌体阻遏蛋白密切相关，如图 7-14 所示）通过一条用线段表示的柔性多肽“绳子”连接起来而形成。整个蛋白质由一个基因编码并以一条连续的肽链合成。两个结构以这种方式共价连接导致蛋白质与其特异的 DNA 序列的亲和力相比任何单独结构的 DNA 亲和力大大提高。一组哺乳动物基因调节蛋白作为这种结构的例子，调控着生长因子，免疫球蛋白和其他参与发育的分子。图中所示的特定例子来自 Oct-1 蛋白（摘自 J. D. Klemm et al., *Cell* 77：21～32，1994）

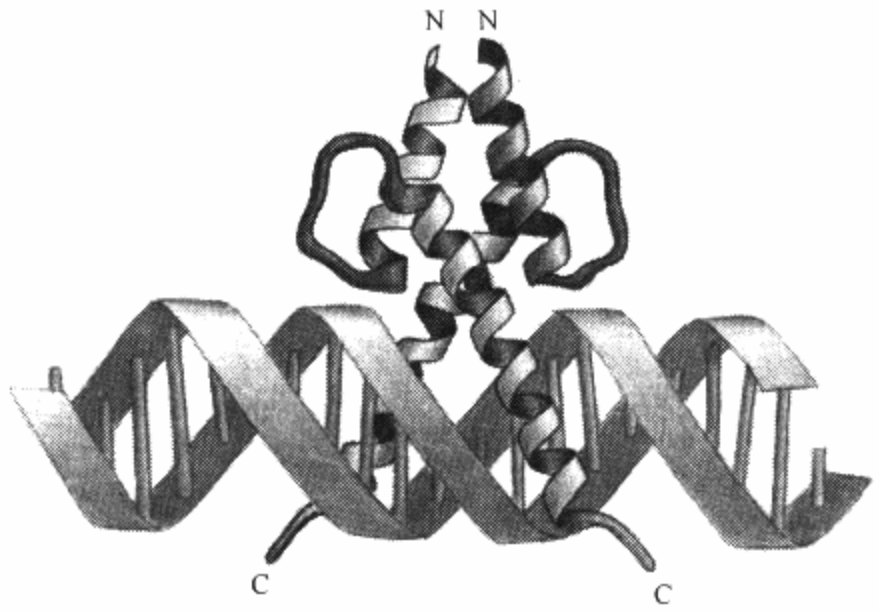


图 7-25　结合 DNA 的一种螺旋-环-螺旋二聚体

两个单体维系在一个四螺旋束中：每个单体贡献两个 α 螺旋，它们由一个蛋白质柔性环连接起来（红色）。特异的 DNA 序列由突出到四螺旋束之外的两个 α 螺旋结合(摘自 A. R. Ferre-D Amare et al. *Nature* 363:38～45,1993)

中的 HLH 基序结合。与亮氨酸拉链蛋白一样，另一个 HLH 蛋白可以相同（形成一个同源二聚体），也可以不同（形成一个异源二聚体）。在两种情况中，从二聚化交界面伸出的两个 α 螺旋与 DNA 之间形成特异的接触。

许多 HLH 蛋白缺少负责结合 DNA 的 α 螺旋延伸。这些截短的蛋白质能与全长的 HLH 蛋白形成异源二聚体，但是这个异源二聚体不能紧密结合 DNA，因为它们只形成所必需的接触中的一半。因此，除了产生活性二聚体以外，异源二聚化也提供了一种方式，控制特定基因调节蛋白（图 7-26）。

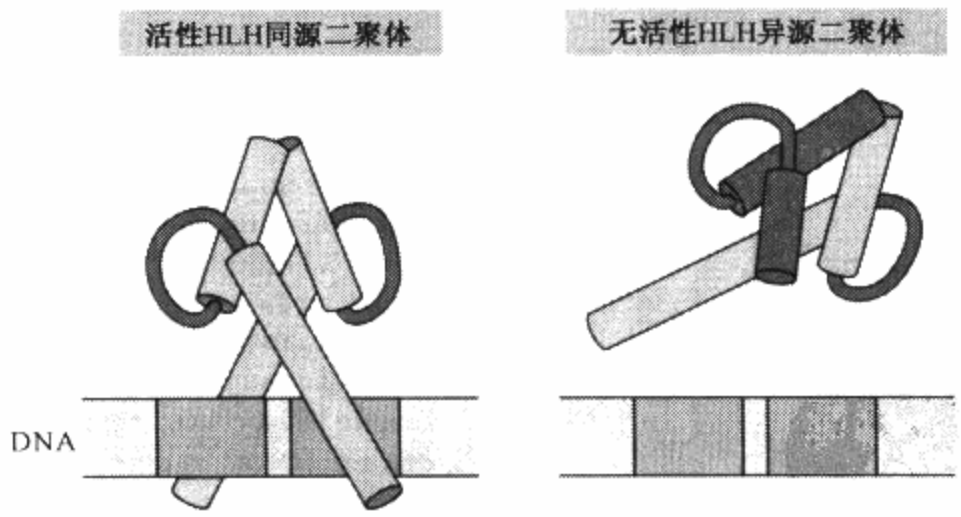


图 7-26　截短的 HLH 蛋白进行的抑制性调节

HLH 基序负责二聚化和 DNA 结合。左边，HLH 同源二聚体识别对称的 DNA 序列。右边，全长 HLH 蛋白（蓝色）与缺少 DNA 结合 α 螺旋的截短的 HLH 蛋白（绿色）结合生成一个异源二聚体，不能牢固结合 DNA。如果过量存在，截短蛋白质分子就会阻断全长 HLH 蛋白的同源二聚体化，因而阻止它结合 DNA

还不能预测出所有基因调节蛋白识别的 DNA 序列

我们已经讨论的不同 DNA-结合基序提供了结构框架，特定的氨基酸侧链从它上面伸出，

与 DNA 中的特异碱基对接触。因此，如果问是否存在一个简单的氨基酸-碱基对识别密码是合理的：例如，是否一个 G-C 碱基对总是与特定的某个氨基酸侧链接触？答案似乎是否定的，尽管某些类型的氨基酸-碱基相互作用看起来比另外的一些频率要高得多（图 7-27）。我们在第 3 章看到，实际上具有任何形状和化学的蛋白质表面可以仅仅由 20 种不同氨基酸形成，而基因调节蛋白利用它们的不同组合形成与特定 DNA 序列精确互补的表面。我们知道相同的碱基对因此由于周围环境不同可以有许多方式进行识别（图 7-28）。然而，分子生物学家正开始深入地了解蛋白质-DNA 识别问题，我们不久就应该能设计出识别任何想要的 DNA 序列的蛋白质。

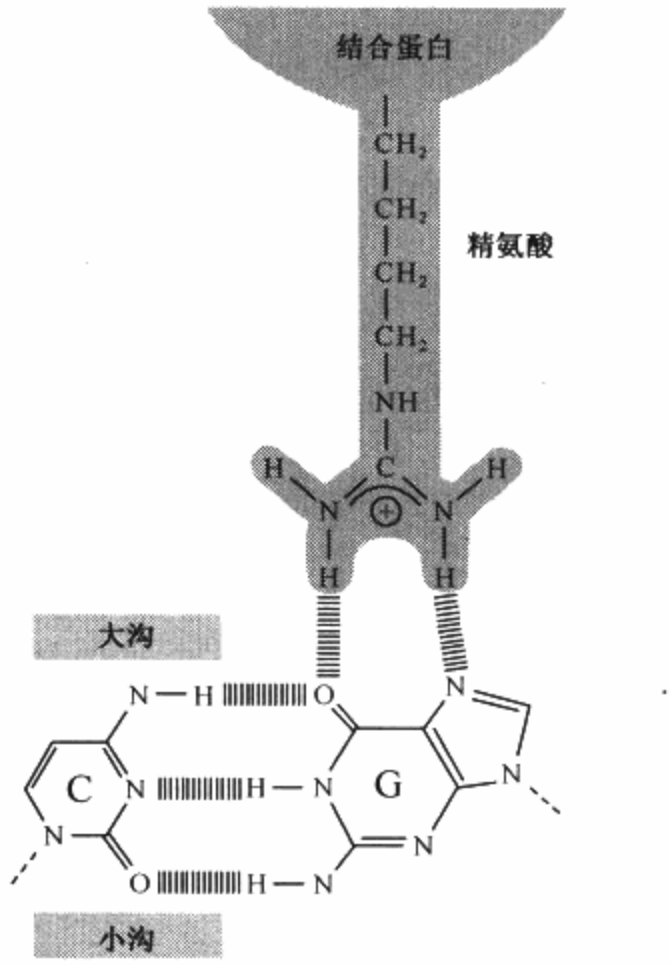


图 7-27　一种最常见的蛋白质-DNA 相互作用

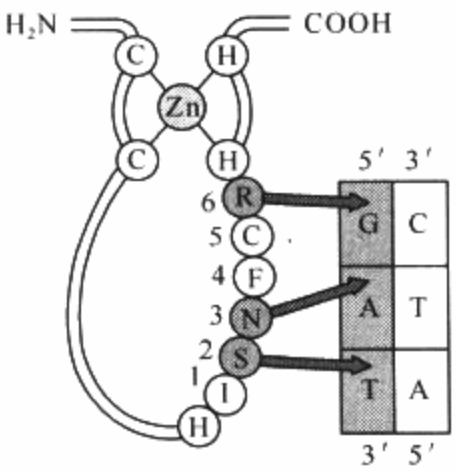
由于其特殊的氢键受体几何学特征（图 7-7），鸟嘌呤能由精氨酸的侧链模糊识别。

图 7-12 中则显示了另一种常见的蛋白质-DNA 相互作用

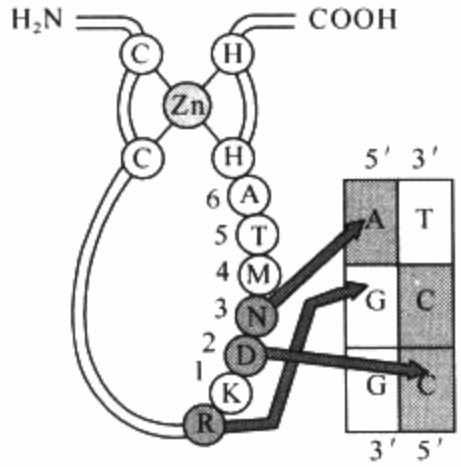
利用凝胶迁移率变动分析能容易地发现序列特异性 DNA 结合蛋白

遗传分析能提供一条发现细菌、酵母和果蝇的基因调节蛋白的途径，但在脊椎动物中就要困难得多。因此，分离脊椎动物基因调节蛋白就必须等到新方法出现。这些方法有许多依赖于在细胞抽提物中 DNA 结合蛋白的发现，这些蛋白质能特异性地识别已知的调控特定基因表达的 DNA 序列。最常见的发现序列特异的 DNA 结合蛋白的方法就是利用一种基于结合蛋白对电场中 DNA 分子移动影响的技术。

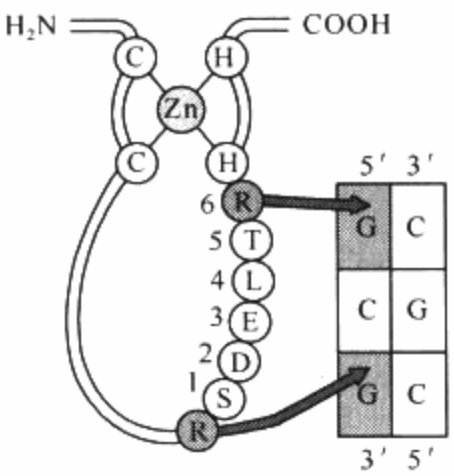
DNA 分子带有大量的负电荷，因而在电场中将迅速向正极移动。当用聚丙烯酰胺凝胶电泳进行分析时，DNA 分子根据其大小被分开，因为较小的分子比大分子更容易穿过细小的凝胶网格。结合到 DNA 分子上的蛋白质分子将使它通过凝胶的速度慢得多；一般来说，结合的蛋白质越大，DNA 分子阻滞就越严重。这种现象为**凝胶迁移率变动**分析提供了基础。这种



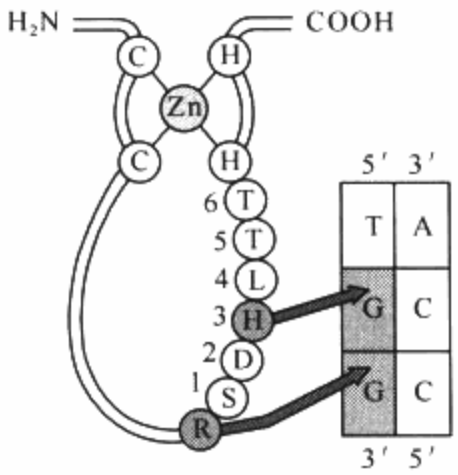
TTK锌指1



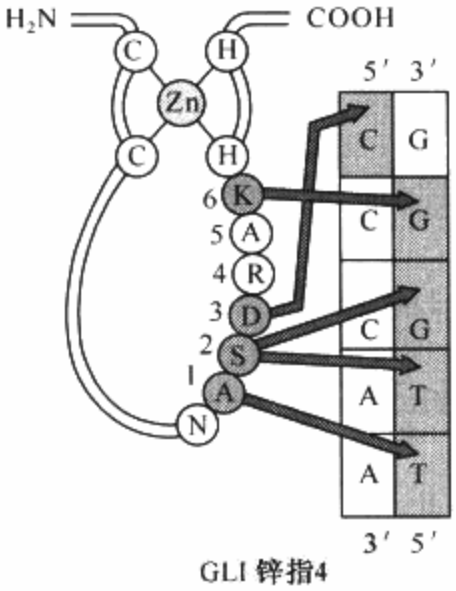
TTK锌指2



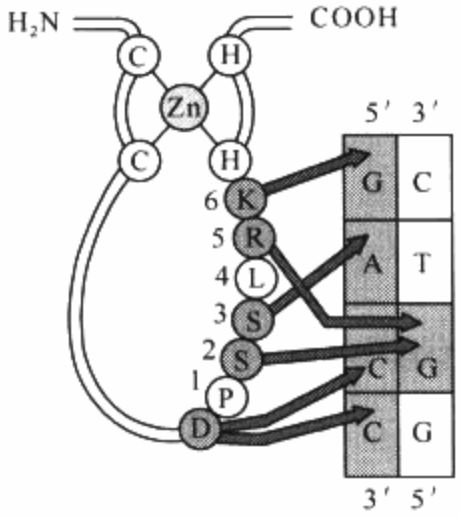
Zif 锌指1



Zif 锌指2



GLI 锌指4



GLI 锌指5

图 7-28　概括不同的 6 锌指与它们的 DNA 识别序列之间序列特异性相互作用

即使全部 6 锌指具有相同的整体结构（图 7-17），但每一个都结合不同的 DNA 序列。标上数字的氨基酸形成识别 DNA 的 α 螺旋（图 7-17 和图 7-18），那些产生序列特异性 DNA 接触的氨基酸用绿色表示。蛋白质接触的残基用橙色表示。虽然精氨酸-嘌呤接触较常见（图 7-27），但嘌呤也能被丝氨酸、组氨酸和赖氨酸识别，如图所示。此外，相同的氨基酸（本例中是丝氨酸）能识别不只一种碱基。所描绘的锌指中有两个来自 TTK 蛋白（一种在果蝇发育过程中起作用的蛋白质）；两个来自图 7-18 中显示的小鼠蛋白（Zif268）；还有两个来自一种人类蛋白（GL1），其畸变形式能引起某些癌症（摘自 C. Branden and J. Tooze, Introduction to Protein Structure, 2nd edn. New York: Garland Publishing, 1999）

方法甚至能容易地测出痕量的序列特异 DNA 结合蛋白。在这种分析中，对一段特定长度和序列的 DNA 端片段（由 DNA 克隆或者化学方法合成得到）进行放射性标记，并与细胞抽提物混合；然后将混合物上样到聚丙烯酰胺凝胶上进行电泳。如果这个 DNA 片段对应一段多个序列特异性蛋白质结合的染色体区域，放射自显影将显示出一系列 DNA 条带，每一条带都不同程度地发生了阻滞，并代表一个不同的 DNA-蛋白质复合物。然后产生凝胶上每一条带的蛋白质彼此间可以通过随后进行的细胞抽提物分级进行分离（图 7-29）。

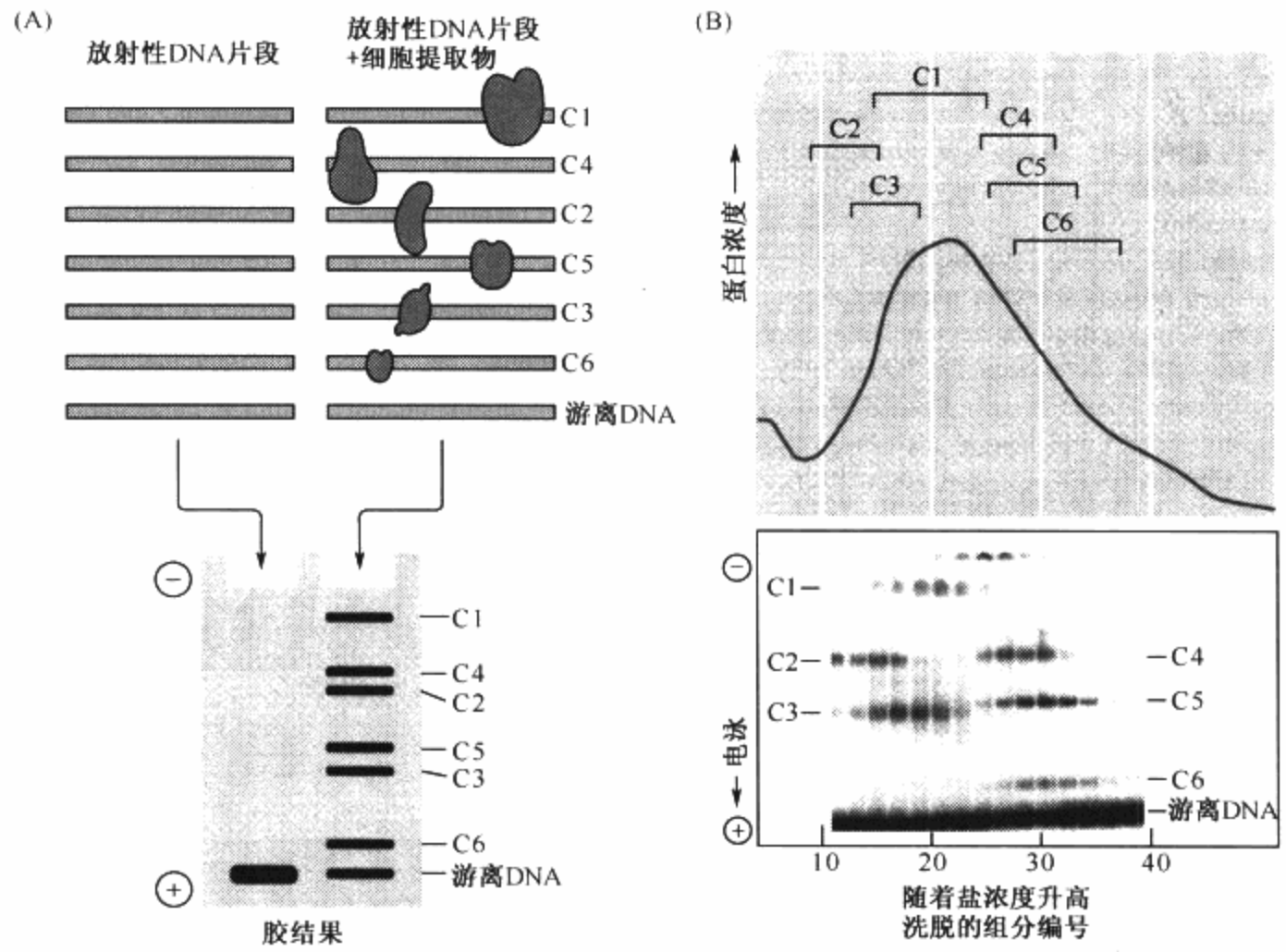


图 7-29　凝胶迁移率变动分析

（A）中概括显示了该分析的原理。本例中将产生抗体的细胞系提取物与放射性 DNA 片段混合，它含有大约 160 个核苷酸的基因调节序列，来自该细胞系的抗体轻链的编码基因。提取物中的蛋白质对 DNA 片段迁移率的影响通过多聚丙烯酰胺凝胶电泳以及随后的放射性自显影进行分析。没有结合蛋白质的 DNA 片段迅速向凝胶底部迁移，而结合了蛋白质的 DNA 片段则受到迟滞；结果找到 6 条迟滞带表明提取物含有 6 种不同的序列特异性 DNA 结合蛋白（用 C1～C6 表示）结合到该 DNA 序列上。（为简单起见，任何结合超过一个蛋白质的 DNA 片段已从图中略去。）在（B）中，利用标准层析技术分级分离（上面），每个分离成分与放射性 DNA 片段混合，然后在聚丙烯酰胺凝胶电泳的一条泳道上进行电泳，并像（A）中那样进行分析[(B)摘自 C. Scheidereit, A. Heguy and R. G. Roeder, *Cell* 51: 783～793, 1987]

DNA 亲和层析使序列特异 DNA 结合蛋白的纯化更容易

一旦某个基因调节蛋白识别的 DNA 序列已经确定，可以应用一种称为 DNA 亲和层析的功能特别强大的纯化方法。用化学方法合成一段正确序列的双链寡聚核苷酸，并连接到不溶性的多孔基质上，如琼脂糖；然后用结合有寡聚核苷酸的基质构建一个柱子，选择性地结合能识别特定 DNA 序列的蛋白质（图 7-30）。利用这种方法只需付出较小的努力就能获得多至 10 000 倍的纯化效率。

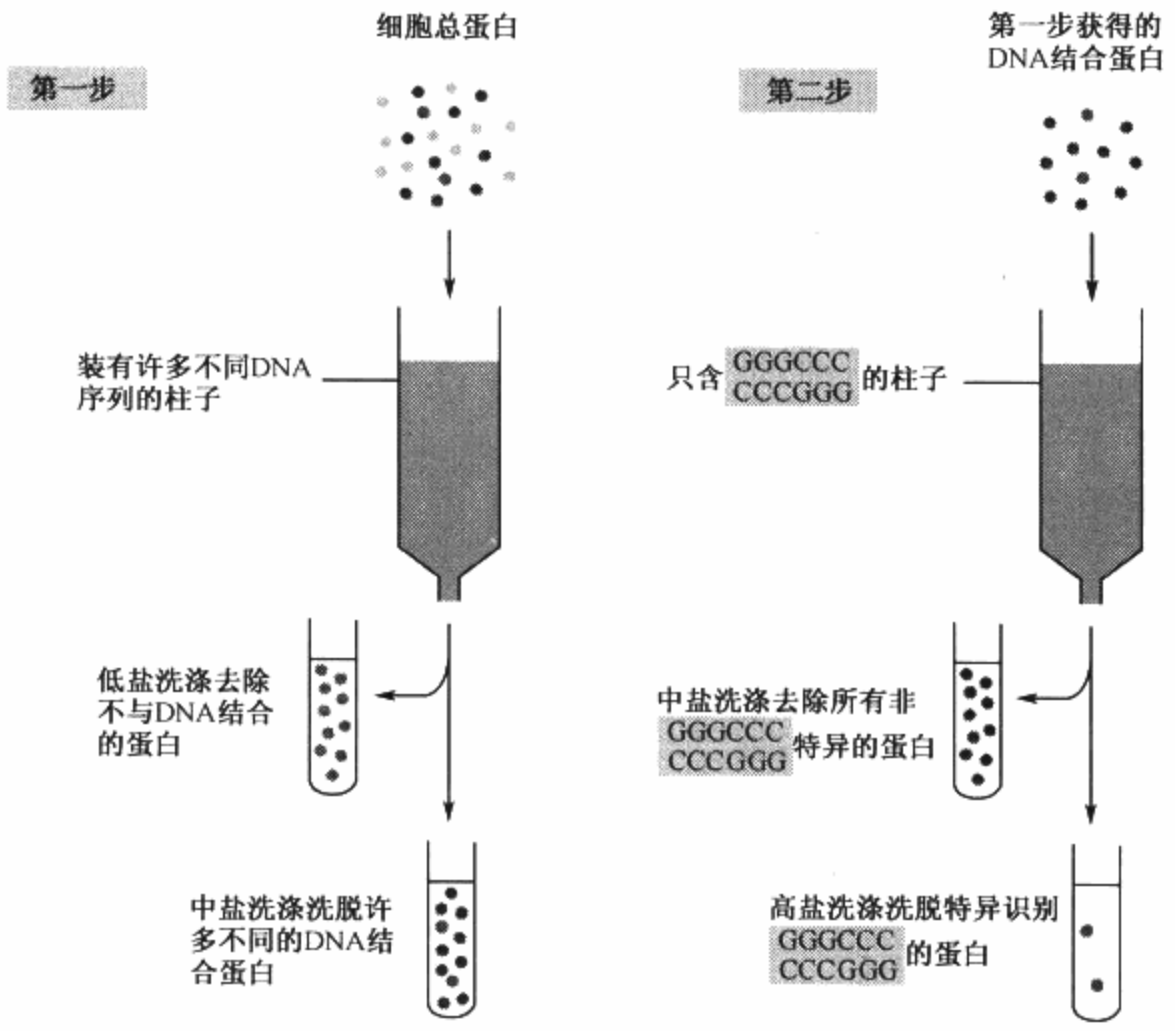


图 7-30　DNA 亲和层析

第一步，在装有大量不同 DNA 序列的柱子中将全部能结合 DNA 的蛋白质与其余细胞蛋白质分离。大多数序列特异的 DNA 结合蛋白对大量 DNA 亲和力弱（非特异性），因而保留在柱子上。这种亲和力大多是由于离子吸引，并且可以用中等浓度的盐溶液将这些蛋白质从 DNA 上洗脱下来。第二步，DNA 结合蛋白混合物经过只含有特定 DNA 序列的柱子。通常情况下，全部 DNA 结合蛋白将会结合到柱子上，大多数是通过非特异相互作用。再用中等浓度盐溶液洗脱这些蛋白质，留在柱子上的都是那些特异性地结合（通常有一种或几种）因而非常紧密地结合特定 DNA 序列的蛋白质。这些留下来的蛋白质可以利用很高盐浓度的溶液从柱子上洗脱

虽然大部分结合特异 DNA 序列的蛋白质在每个较高等的真核细胞中存在的拷贝数达到几千个（而且一般只代表总共 50 000 种细胞总蛋白质的一部分），为了利用质谱或其他方式（讨论见第 8 章）获得蛋白质的部分氨基酸序列，通常可以利用亲和层析分离足够纯化的蛋白质（讨论见第 8 章）。如果生物完整的基因组序列已知，这部分氨基酸序列就可以用来鉴定基因。这个基因提供了该蛋白质完整的氨基酸序列，而有关外显子和内含子边界的任何不确定性都能通过分析这个基因产生的 mRNA 得到解决，见第 8 章描述。这个基因也提供了通过基因工程技术无限量合成该蛋白质的方式（讨论见第8 章）。

可以确定出一个基因调节蛋白识别的 DNA 序列

有些基因调节蛋白在它们结合的 DNA 序列已知以前就发现了。例如，许多果蝇同源异型结构域蛋白是通过分离翅膀发育改变了的突变体而发现的。这使得编码这些蛋白质的基因能够鉴定出来，然后这些蛋白质可以在培养细胞中过量表达并容易纯化得到。确定一种基因调节蛋白识别的 DNA 序列的方法之一就是利用纯化的蛋白质从大量具有不同序列的短核苷

酸库中筛选出与它紧密结合的那些序列。经过多轮筛选，可以确定出紧密结合的 DNA 的核苷酸序列，继而可以设计出针对这个基因调节蛋白的共有 DNA 识别序列（图 7-31）。电脑可以利用共有序列来搜寻基因组序列，因而鉴定出转录可能受到这个感兴趣的基因调节蛋白调控的候选基因。然而这种策略并不是万无一失的。例如，许多生物产生一组密切相关的基因调节蛋白，识别非常类似的 DNA 序列，这种方法就不能解决它们的问题。在大多数情况下，通过基因组搜寻获得的基因调节蛋白作用位点的估计必须利用更直接的方法，比如下一部分要描述的方法，来检验。

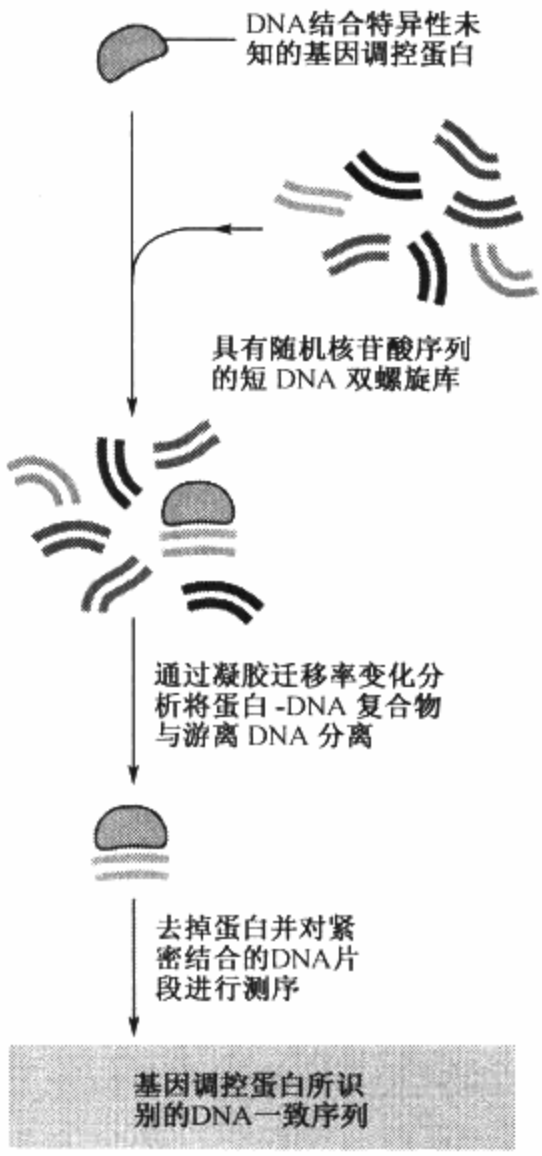


图 7-31　确定基因调节蛋白识别的 DNA 序列的方法

纯化的基因调节蛋白与数百种不同的短 DNA 片段混合，这些片段每个都有不同的核苷酸序列。一组这样的 DNA 片段可以利用程序化 DNA 合成仪合成，它是化学合成所需要的 DNA 序列的仪器（讨论见第 8 章）。例如，一段 11 核苷酸的 DNA 片段就有 4^{11} 或将近 420 万种可能的核苷酸序列。然后紧密结合基因调节蛋白的双链 DNA 序列与不能结合的序列分离。一种实现这种分离的方法就是通过凝胶迁移率变动分析，如图 7-29 所述。DNA-蛋白质复合物与游离 DNA 分离以后将 DNA 片段从蛋白质中除去，并重复进行多次相同的筛选流程。经过多轮筛选仍然裸露的 DNA 片段的核苷酸序列可以进行确定，并得到一个共有 DNA 识别序列

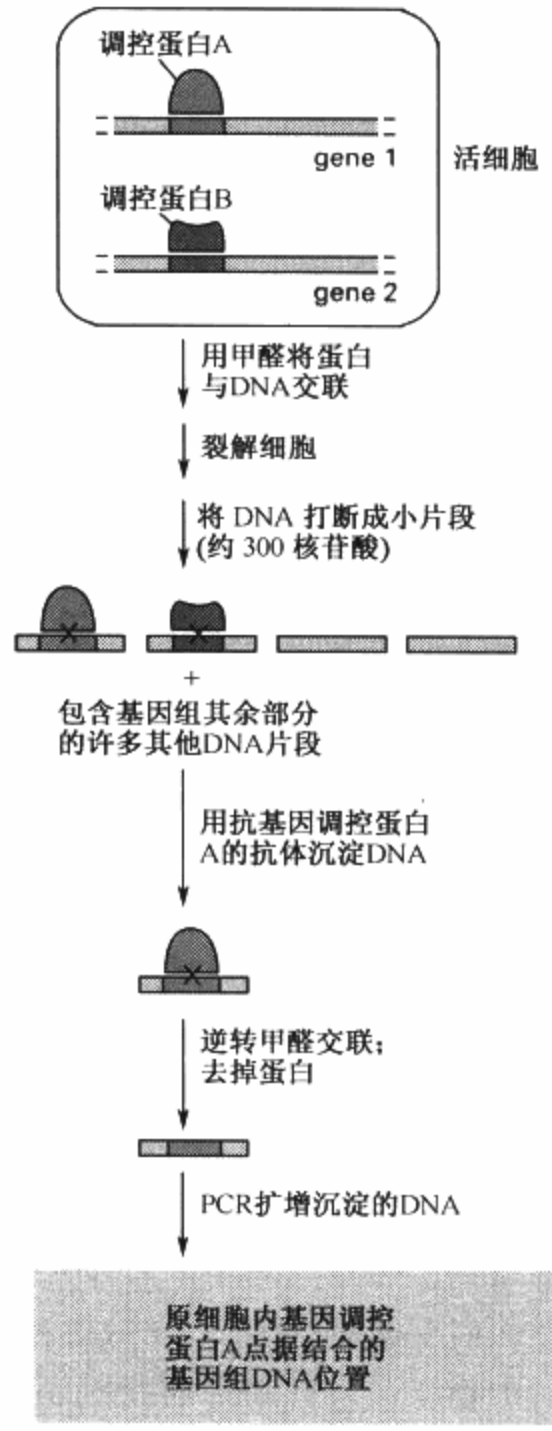


图 7-32　染色质免疫沉淀

该方法能鉴定出体内基因调节蛋白占据的基因组中的位点。利用聚合酶链反应（PCR）扩增 DNA 将在第 8 章描述。将片段混合物与 DNA 微阵列杂交可以确定发生沉淀的扩增 DNA 片段，这些也将在第 8 章描述

染色质免疫沉淀技术可以在活细胞中鉴定出基因调节蛋白结合的DNA位点

一般来说，给定的某个基因调节蛋白不能任何时候都占有基因组内所有潜在的DNA结合位点。在有些情况下，可能这个蛋白质没有合成，因而在细胞内不存在；或者，例如，它可能存在但是为了在活细胞中有效地结合DNA，已经与其他蛋白质形成了异源二聚体；或者它可能排除在核外，直到从细胞所处环境中接受到合适的信号。一种称为**染色质免疫沉淀**的方法能凭经验确定出给定基因调节蛋白在特定的一组条件下占有的DNA位点（图7-32）。在活细胞内将蛋白质DNA共价交联起来，然后裂解细胞，将DNA用机械力打断成小片段。由于交联时蛋白质接近DNA，所以接着用针对给定基因调节蛋白的抗体可以纯化共价交联在这个基因调节蛋白上的DNA。使用这种方法，可以确定出原先细胞内基因调节蛋白占有的DNA位点。

这种方法也常用来鉴定由各种类型的修饰组蛋白包装起来的基因组上的位点（图4-35）。在这种情况下，应用了对特定组蛋白修饰特异的抗体。

小结

基因调节蛋白识别具有确定序列的双螺旋DNA短片段，因而决定了细胞中数千个基因的哪一个将被转录。在各种生物中已鉴定出数千种基因调节蛋白。虽然这些蛋白质每一种都有其独特的特征，可是大多数都以同源二聚体或异源二聚体形式结合DNA，并通过少量几种结构基序中的一种识别DNA。常见的基序包括螺旋-转角-螺旋、同源异型结构域、亮氨酸拉链、螺旋-环-螺旋和各种类型的锌指。折叠成基序的精确氨基酸序列决定了它识别的DNA序列。异源二聚化增加了基因调节蛋白能够识别的DNA序列范围。利用基因调节蛋白的DNA序列特异性，许多功能强大的技术可以用来鉴定和分离这些蛋白质，编码它们的基因，它们识别的DNA序列，以及它们调节的基因。

基因开关如何工作

在前面的部分中，我们描述了基因开关的基本组分——基因调节蛋白和这些蛋白质识别的特异DNA序列。现在我们要讨论这些组分如何运行，从而响应不同的信号将基因打开或关闭。

就在40年前，认为基因能够打开和关闭还是全新的观点。这种观点是一个主要的进步，它最初来源于研究大肠杆菌如何适应它们生长培养基中条件的改变。在λ噬菌体中进行的类似实验得到许多相同的结论，并帮助确立了其内在的机制。许多相同的原理也适用于真核细胞。然而由于在高等生物中的调控极其复杂，以及它们的DNA包装成染色质，所以产生了特殊的困难以及某些调控的新机制，这些我们将会看到。这里我们首先讨论最简单的例子，细菌中一种应答单个信号的开关。

色氨酸阻遏蛋白是在细菌中将基因打开或关闭的一种简单开关

单细胞生物大肠杆菌的染色体由一条长约4.6×10^6个核苷酸对的环状DNA分子组成。这条DNA编码了将近4300个蛋白质，虽然任何一次只合成其中的一部分。许多这些蛋白质的表达根据环境中可获取的食物情况受到调控。这一点可以通过编码合成色氨酸的5个酶的大肠杆菌基因加以说明。这些基因排列成单个**操纵子**；也就是说它们在染色体上彼此相邻，

并从单个启动子上以一条长 mRNA 分子转录出来（图 7-33）。但是当生长培养基中存在色氨酸并进入细胞内时（例如，该细菌位于刚刚采食蛋白质后的哺乳动物肠道中），细胞就不再需要这些酶，并将它们的合成关闭。

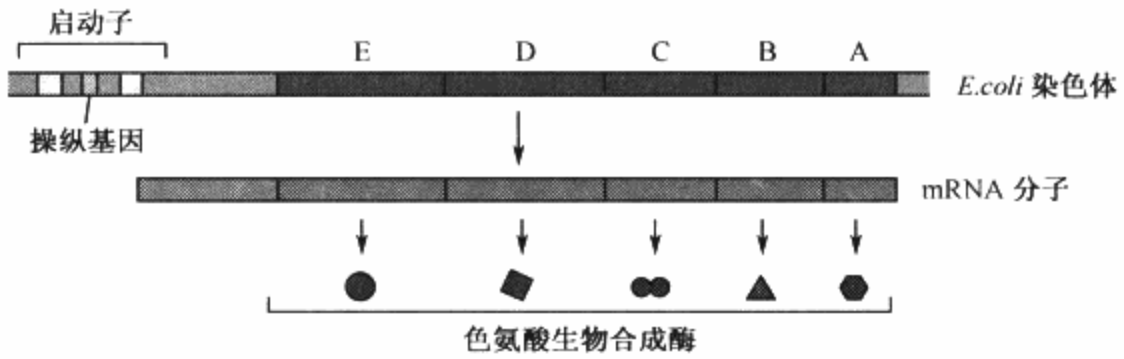


图 7-33　大肠杆菌中成簇的编码合成色氨酸的酶的基因

这 5 个基因作为一条 mRNA 分子转录，该特征使它们的表达能够协调控制。以单个 mRNA 分子转录的成簇基因在细菌中经常见到，每一个这样的基因簇称为一个操纵子

有关这个开关的分子基础已有了大量详细的了解。已在第 6 章描述过，启动子是一段特殊的 DNA 序列，指导 RNA 聚合酶结合到 DNA 上，打开 DNA 双螺旋，并开始合成一条 RNA 分子。在指导色氨酸生物合成基因转录的启动子内，存在一个称为**操纵基因**的调控元件（图 7-33）。它只是调节 DNA 中由阻遏蛋白识别的有确定核苷酸序列的一小段区域，在这个例子中，该阻遏蛋白称为**色氨酸阻遏蛋白**，是螺旋-转角-螺旋家族的成员之一（图 7-14）。细胞这样安排启动子和操纵基因，当色氨酸阻遏蛋白占据操纵基因时，就会阻断 RNA 聚合酶进入启动子，因而阻止色氨酸合成酶的表达（图 7-34）。

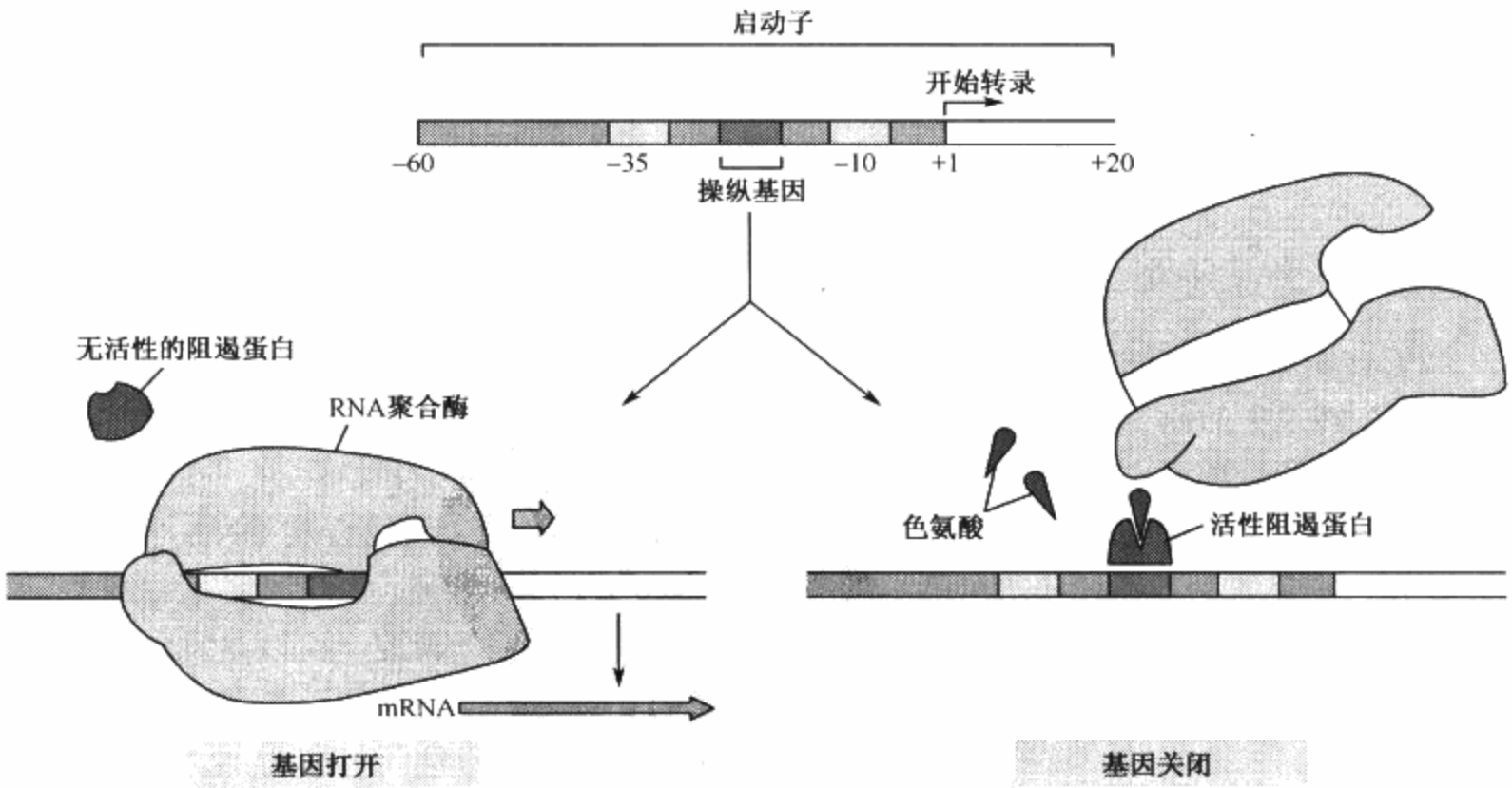


图 7-34　色氨酸基因的打开与关闭

如果细胞内的色氨酸水平较低，RNA 聚合酶就结合到启动子上转录色氨酸操纵子的 5 个基因。然而，如果色氨酸水平较高，色氨酸阻遏蛋白活化后结合到操纵基因上，在这里它阻断 RNA 聚合酶结合到启动子上。无论什么时候细胞内色氨酸水平下降，阻遏蛋白就释放其色氨酸并失活，从而允许聚合酶开始转录这些基因。启动子包括两个重要的 DNA 序列信息区域，用黄色突出显示的－35 区和－10 区（图 6-12）

基因表达的阻断受一种精巧的方式调控：为了结合到它的操纵基因 DNA 上，阻遏蛋白必须结合有两分子的色氨酸。如图 7-35 所示，色氨酸的结合使阻遏蛋白的螺旋-转角-螺旋基序发生倾斜，这样它就能正确地提供给 DNA 大沟结合；如果没有色氨酸，基序就向内摆动，该蛋白质就不能结合到操纵基因上。因此色氨酸阻遏蛋白和操纵基因形成一个简单的装置，根据可获得的游离色氨酸的情况，打开或关闭色氨酸生物合成酶的合成。由于蛋白质活化的 DNA 结合形式起关闭基因的作用，所以这种基因调控模式称为**负调控**，而以这种方式发挥功能的基因调节蛋白称为转录阻遏蛋白或基因阻遏蛋白。

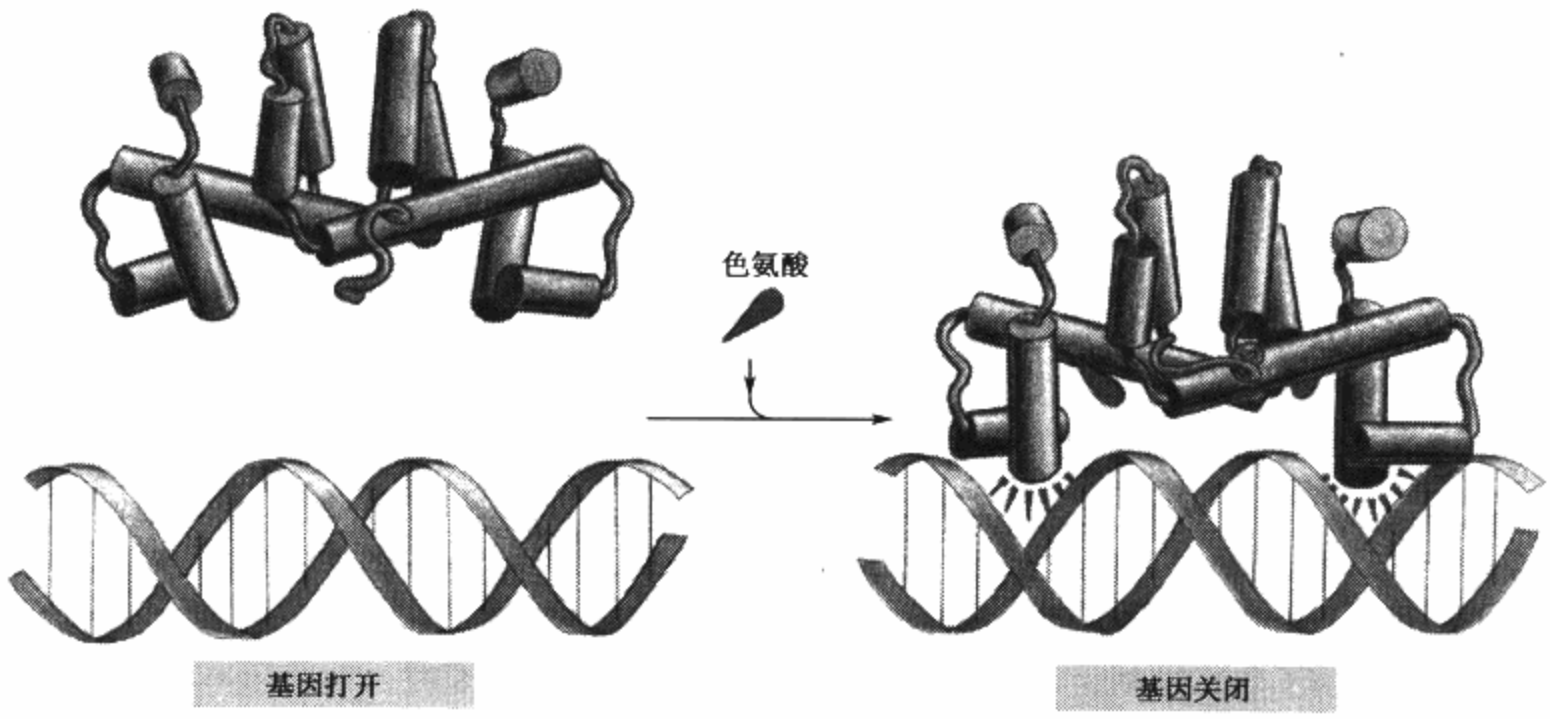


图 7-35　色氨酸结合色氨酸阻遏蛋白改变了阻遏蛋白的构象

构象改变使该基因调节蛋白能够紧密结合特异的 DNA 序列（操纵基因），因而阻断编码合成色氨酸所需酶的基因的转录（trp 操纵子）。图中说明了整个细菌螺旋-转角-螺旋蛋白的三维结构，分别在有色氨酸结合和无色氨酸结合的情况下利用 X 射线衍射得到。色氨酸结合增加了同源二聚体中两个识别螺旋的距离，因而使阻遏蛋白与操纵基因紧挨着结合起来

（摘自 R. Zhang et al., *Nature* 327：591～597，1987）

转录激活蛋白将基因打开

我们在第 6 章看到纯化的大肠杆菌 RNA 聚合酶（包括 σ 亚基）能结合到启动子上并能够起始 DNA 转录。然而有些细菌启动子自身只发挥一小部分的功能，或者是因为它们不能很好地被 RNA 聚合酶识别，或者因为聚合酶难以打开 DNA 螺旋并开始转录。在这两种情况下，这些功能较弱的启动子可以由基因调节蛋白挽救，这些蛋白质结合到 DNA 上邻近的位点，并以一种显著增加转录起始可能性的方式与 RNA 聚合酶接触。由于这种蛋白质以活性的 DNA 结合形式将基因打开，所以将这种基因调控模式称为**正调控**，而以这种方式发挥功能的基因调节蛋白称为转录激活因子或者基因活化蛋白。在有些情况下，细菌基因活化蛋白为 RNA 聚合酶提供额外的接触表面，从而帮助 RNA 聚合酶结合到启动子上。在其他情况下，它们可能通过稳定某种过渡状态促使聚合酶由起始 DNA 结合构象向活跃转录形式转变。

正如在转录阻遏蛋白进行的负调控那样，转录激活蛋白能作为一种简单开启-关闭开关的成分之一发挥作用。例如，当没有葡萄糖，细菌优先利用的碳源可以利用时，细菌活化蛋白质 CAP（分解代谢物激活蛋白）激活那些能使大肠杆菌利用替代碳源的基因。葡萄糖水平下降会诱发结合 CAP 蛋白的细胞内信号分子环 AMP 增加，从而使该蛋白质能在目标启动子附

近结合其特异的 DNA 序列，并因而打开适当的基因。通过这种方式，根据细胞中环 AMP 水平是高还是低，决定将目标基因是打开还是关闭。图 7-36 概括了可用来调节基因的不同正调控和负调控方式。

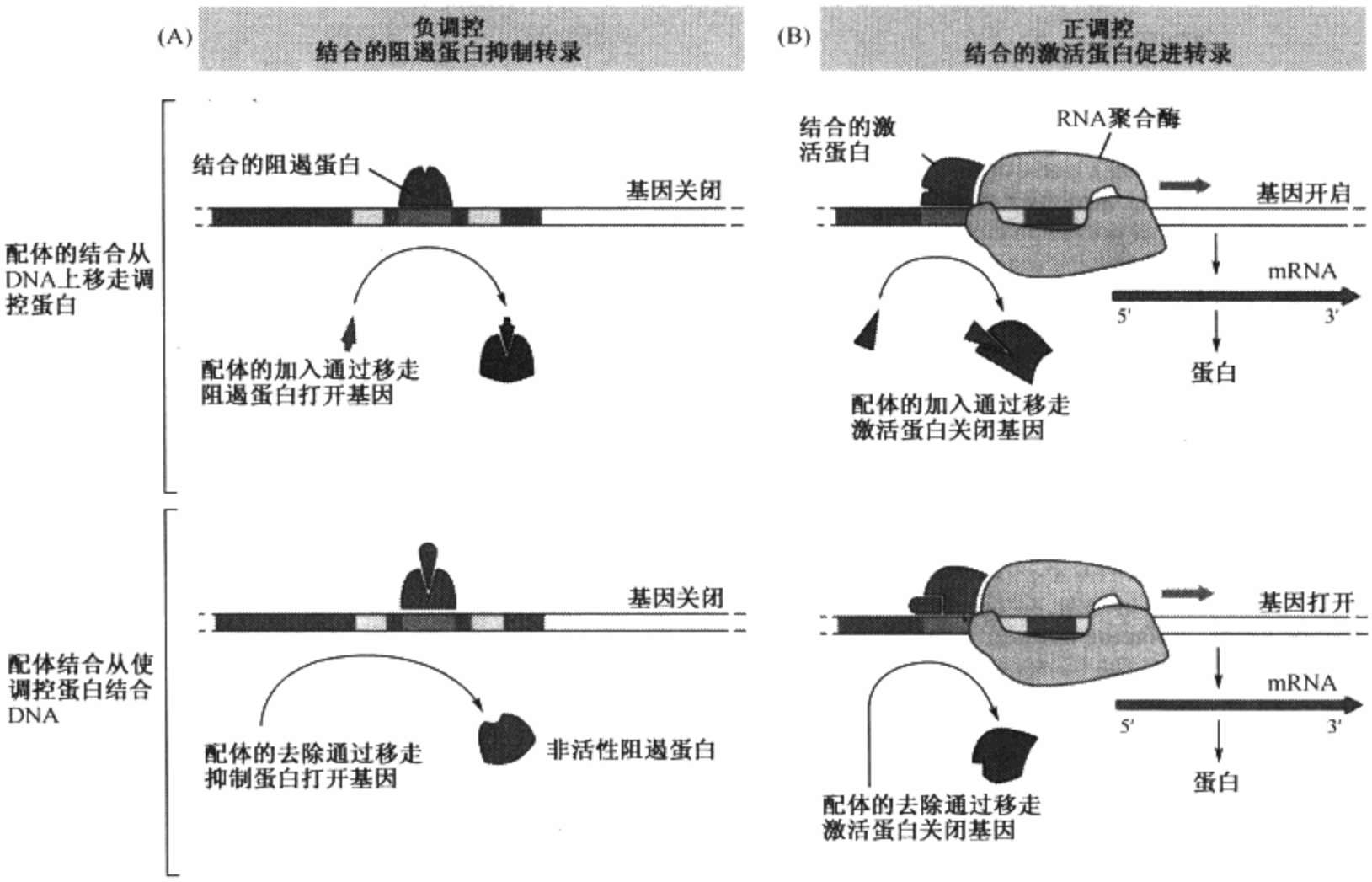


图 7-36　概括原核生物中特定基因调节蛋白调控基因转录的机制

（A）负调控；（B）正调控。注意“诱导”配体的加入能打开基因，或者是通过从 DNA 移走基因阻遏蛋白（左上图）或者是引起基因激活蛋白的结合（右下图）。同样，“抑制”配体的加入能关闭基因，或者是通过从 DNA 上移走基因激活蛋白（右上图），或者是引起基因阻遏蛋白的结合（左下图）

转录激活蛋白与转录阻遏蛋白的结构在许多方面都相似。例如，色氨酸阻遏蛋白和转录激活蛋白 CAP 都利用了一个螺旋-转角-螺旋基序（图 7-14），并且为了结合 DNA，两者都需要一个小辅助因子。实际上，有些细菌蛋白质（包括 CAP 和 λ 噬菌体阻遏蛋白）既能作为激活蛋白，又能作为阻遏蛋白，这取决于它们识别 DNA 序列相对于启动子的确切位置：如果蛋白质结合位点与启动子重叠，就不能结合聚合酶，该蛋白质就作为一个阻遏蛋白（图 7-37）。

一个转录激活蛋白和一个转录阻遏蛋白调控 lac 操纵子

更复杂的基因开关类型将正调控与负调控组合起来。例如，与 trp 操纵子不同，大肠杆菌中的 lac 操纵子由 lac 阻遏蛋白和 CAP 分别进行负和正的转录调控。lac 操纵子编码的蛋白质用来转运乳糖到细胞内并将其降解。我们已看到，CAP 使细菌在缺少葡萄糖时能利用可替代碳源（如乳糖）。然而，如果不存在乳糖时，CAP 也诱导 lac 操纵子表达的话，就会很浪费，而乳糖阻遏蛋白确保了在乳糖缺乏时将 lac 操纵子关闭。这种安排使 lac 操纵子的调控区域能够应答并整合两种不同的信号，这样只有当遇到这两种情况时，即必须存在乳糖，不存在葡萄糖时，操纵子才高度表达。任何其他 3 种可能的信号组合都将这一串基因保持在关闭状态（图 7-38）。

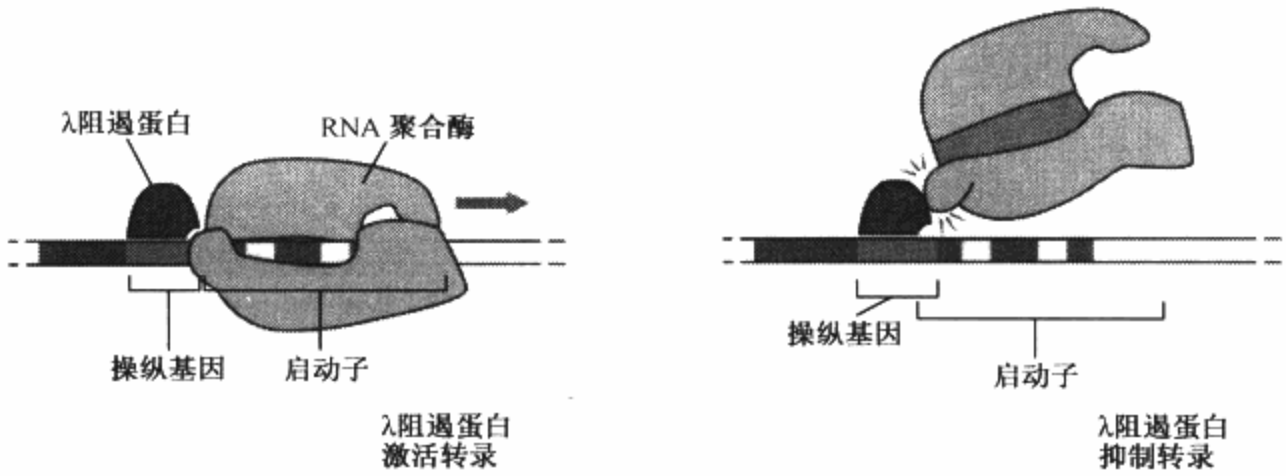
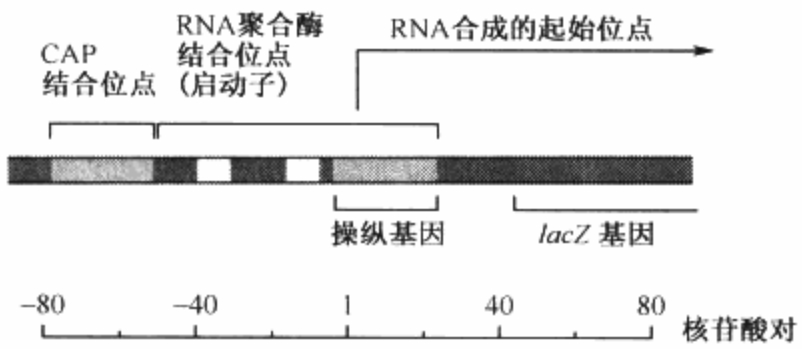


图 7-37　根据在 DNA 中结合位点上的精确定位，有些细菌基因调节蛋白既能作为转录激活蛋白又能作为阻遏蛋白

一个实例就是 λ 噬菌体阻遏蛋白，对于有些基因，该蛋白质提供有利于 RNA 聚合酶（上面）的接触，因而作为转录激活蛋白起作用。对于另一些基因（底部），操纵基因离启动子距离减少一个碱基对，因而阻遏蛋白现在与 RNA 聚合酶竞争结合 DNA，而不是帮助它。如图 7-14 所示，λ 阻遏蛋白通过螺旋-转角-螺旋基序识别它的操纵基因



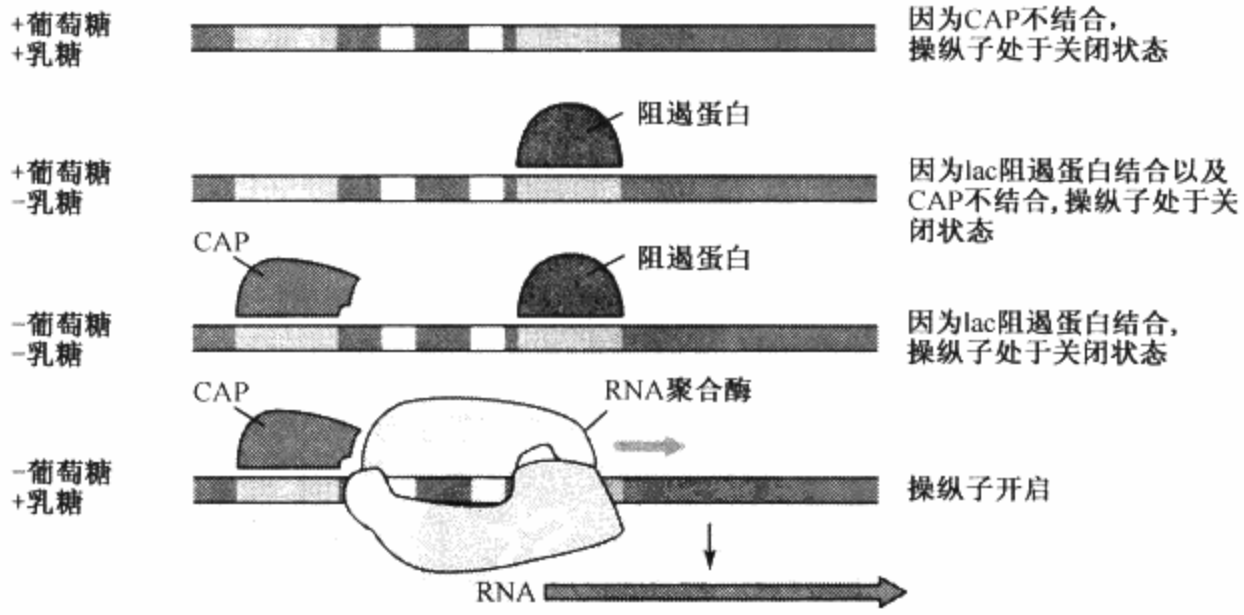


图 7-38　lac 操纵子的双重调控

葡萄糖和乳糖水平通过影响 lac 阻遏蛋白和 CAP 来调控 lac 操纵子转录的起始。增加乳糖就会增加异乳糖浓度，它结合阻遏蛋白并将其从 DNA 移开。增加葡萄糖就会减少环 AMP 的浓度；由于环 AMP 不再结合 CAP，该基因激活蛋白就从 DNA 上解离，关闭了操纵子。如图 7-11 所示，已知 CAP 结合时向 DNA 中引入弯曲；为了简便起见，这里未显示出这种弯曲。lacZ，lac 操纵子的第一个基因，编码了 β-半乳糖苷酶，它将二糖乳糖断裂为半乳糖和葡萄糖。该图概括了 lac 最本质的特征，但实际上情况更加复杂。一方面，DNA 上有几个 lac 阻遏蛋白结合位点。虽然图中所说明的结合位点发挥了最大的作用，但其他的也是完全阻遏所需的。此外，lac 操纵子表达从未完全关闭。还需要少量的 β-半乳糖苷酶将半乳糖转化为异乳糖，因而在生长培养基中加入乳糖时会使 lac 阻遏蛋白失活

50 多年前，这种基因开关的简单逻辑首先引起了生物学家的注意。正如上面所解释的，这个开关的分子基础通过结合遗传学和生物化学才解释出来，从而最先提供了对基因表达如何调控的深刻认识。虽然在高等生物中也利用相同的基本策略调控基因表达，但是所利用的基因开关常常要复杂得多。

真核细胞中转录调控很复杂

调节 lac 操纵子的双信号开关机制精巧而又简单。然而，难以想像它是怎样变成一个复杂系统，使几十种信号能够调控操纵子的转录：在启动子附近没有足够的空间装进足够多的调节 DNA 序列。那么真核生物是如何克服这些限制并形成它们更复杂的基因开关呢?

真核生物中的转录调控与通常在细菌中发现的转录调控在三个重要方面存在差别。

首先，真核生物利用基因调节蛋白，即便它们结合的 DNA 远离它们施加影响的启动子达成千上万个核苷酸对，它们也能发挥作用，这就意味着单个启动子可以受到散布在 DNA 上的几乎无限数量的调节序列的调控。

其次，我们在最后一章会看到，转录所有蛋白质编码基因的真核生物 RNA 聚合酶 Ⅱ 不能独立起始转录。它需要一组称为通用转录因子的蛋白质，在转录开始进行之前，它们必须在启动子上组装起来（“通用”这个词是指这些蛋白质在所有 RNA 聚合酶 Ⅱ 转录的启动子上组装这个事实；在这个方面它们与基因调节蛋白不同，后者只在特定基因上发挥功能）。从理论上讲，这个组装过程提供了多个步骤，这样可以响应调节信号，在每一步加速或减慢转录起始速率，并且许多真核生物基因调节蛋白对这些步骤施加了影响。

第三，真核生物 DNA 包装成染色质为调控提供了很多细菌所不具备的机会。

既然第 6 章已讨论过 RNA 聚合酶 Ⅱ 的通用转录因子，这里我们主要集中在第一和第三种特征，以及它们怎样选择性地用于调控真核基因表达。

真核生物基因调节蛋白从距离调控基因表达

与细菌相似，真核生物利用基因调节蛋白（活化蛋白和阻遏蛋白）调节它们的基因表达，但其方式又有些不同。真核基因激活蛋白结合的 DNA 位点最初称为**增强子**，因为它们的存在显著地“增强了”或者加速了转录速率。1979 年发现这些激活蛋白能远离启动子几千个核苷酸结合时，很让人吃惊。此外，真核激活蛋白结合到真核基因上游或下游时都能影响它的转录。增强子序列与结合它们的蛋白质怎么能相距这么远的距离发挥作用呢？它们与启动子又是如何通讯的呢?

已经提出许多“远距离作用”的模型，但是其中最简单的似乎适用于大多数情况。增强子和启动子之间的 DNA 形成一个突环，使结合在增强子上的激活蛋白能与结合到启动子上的蛋白质（RNA 聚合酶，通用转录因子之一，或其他蛋白质）接触（图 6-19）。因此这段 DNA 作为绳子，帮助增强子上结合的蛋白质与即使相距远达几千个核苷酸对的启动子上结合的蛋白复合物相互作用（图 7-39）。这种现象也在细菌中发生，尽管较不常见，且在短得多的 DNA 上发生（图 7-40）。

真核基因调控区由一个启动子和调节 DNA 序列组成

由于结合到远离启动子的 DNA 上时，真核基因调节蛋白能够调控转录，所以控制基因表达的 DNA 序列常常占据很长一段 DNA。我们将用**基因调控区**这个词指参与调节一个基因转录的全部 DNA（包括**启动子**），在这里通用转录因子和聚合酶进行组装，全部**调节序列**结

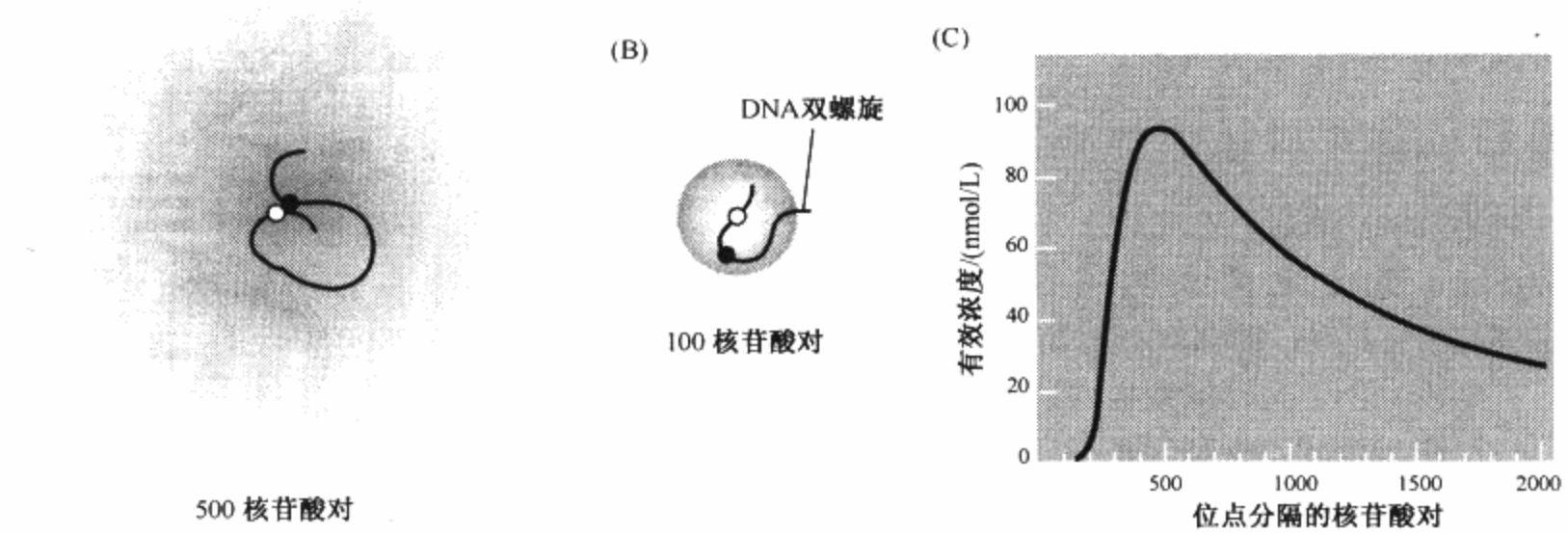


图 7-39　两个蛋白质与 DNA 双螺旋上分开的位点结合能大大增加它们相互作用的可能性

（A）两个蛋白质通过中间长 500 个核苷酸对的 DNA 环连接起来就增加了它们碰撞的频率。蓝色着色的程度反映了红色蛋白质相对于白色蛋白质将定位在空间每个位置的可能性。（B）DNA 具有柔性，序列形成平滑的坡度，一般约每隔 200 核苷酸对形成一次 90 度弯曲（弯曲的转角）。因此，当两个蛋白质仅仅相隔 100 个核苷酸对连接起来，它们的接触就比较受限制。在这些情况下，当分隔两种蛋白质的核苷酸对数目为 10 的倍数时，就能促进蛋白质相互作用。这些核苷酸将两个蛋白质置于蛋白质螺旋（大约每一圈有 10 个核苷酸）的同一侧，并因而位于 DNA 环内，在这里它们可以彼此接触得非常好。（C）在白色蛋白质结合的位点上红色蛋白质理论上的有效浓度，作为它们分隔距离的函数[(C)经 Gregory Bellomy 许可，摘自 M. C. Mossing and M. T. Record, *Science* 233：889～892；1986. © AAAS]

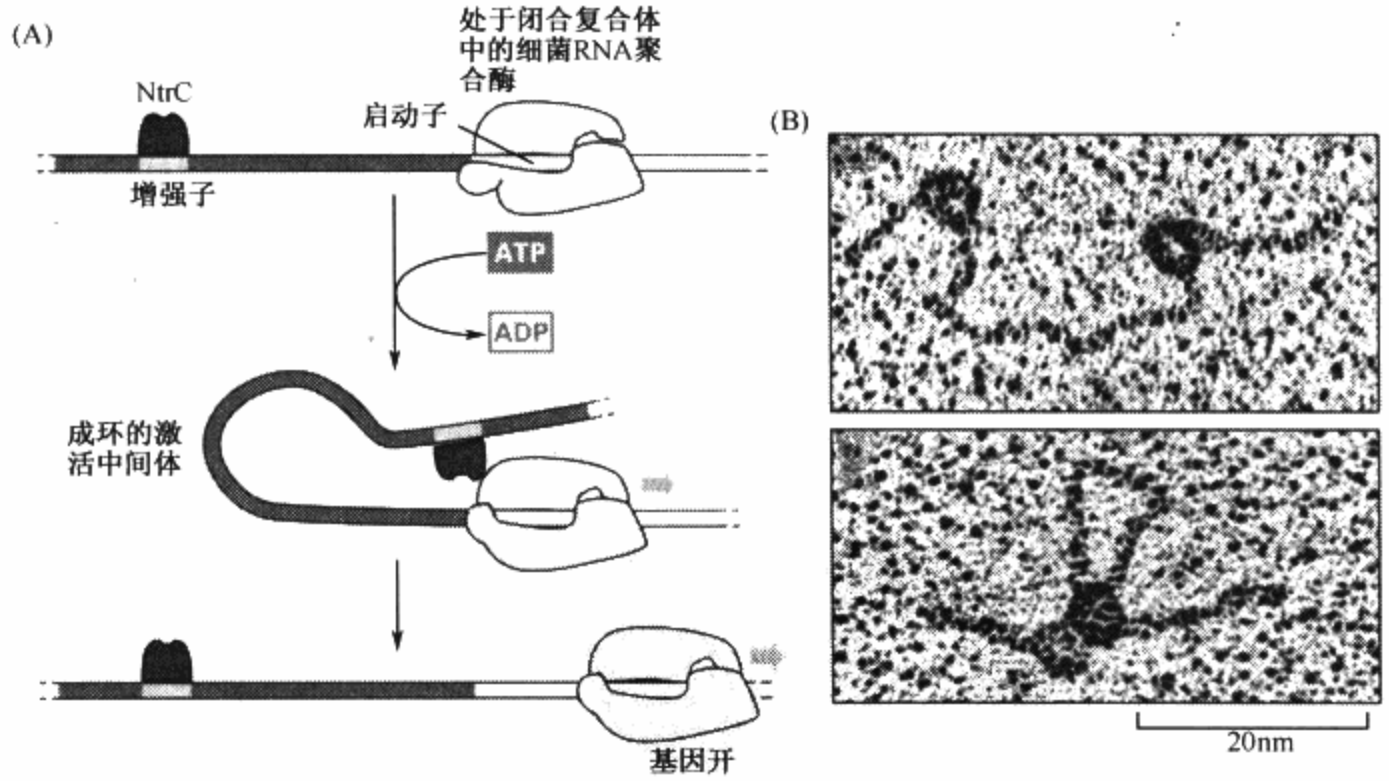


图 7-40　远距离的基因活化

（A）Ntrc 是一种细菌基因调节蛋白，它通过促进 RNA 聚合酶从最初的启动子结合状态向形成起始复合物状态的过渡来激活转录（讨论见第 6 章）。如图所示，Ntrc 刺激的转变需要 ATP 水解产生的能量，尽管这种要求对于细菌转录起始来讲比较少见。（B）电镜中可观察到 Ntrc 与 RNA 聚合酶相互作用，位于它们之间的 DNA 形成环。虽然 DNA 成环引起的转录激活在细菌中少见，但对于真核生物基因调节蛋白来讲则比较典型[(B) 经 Harrison Echols and Sydney Kustu 许可]

合上基因调节蛋白，来调节启动子上组装过程的速率（图 7-41）。在高等生物中，发现某个基因的调节序列长达 50 000 个核苷酸对也并不少见。虽然这段 DNA 大部分作为“间隔”序列，不能由基因调节蛋白所识别，但是这种间隔 DNA 可能通过提供 DNA 结合蛋白之间通讯所需的灵活性来促进转录。还有重要的一点是记住与其他真核染色体区域一样，大部分基因调控区的 DNA 都包装成核小体和更高级的染色质形式，因而压缩了它的长度。

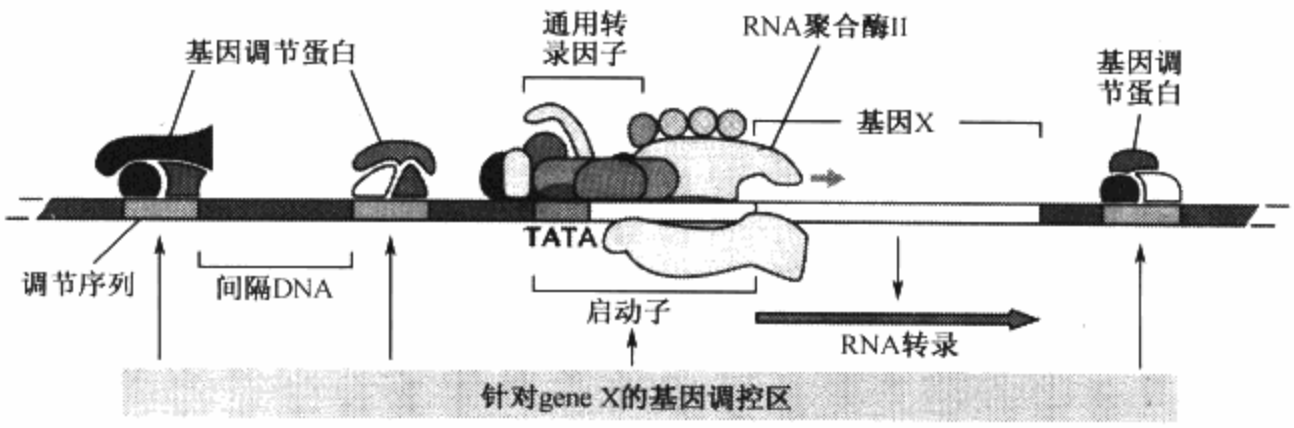


图 7-41　典型真核基因的基因调控区

启动子是通用转录因子和聚合酶装配的 DNA 序列（图 6-16）。调节序列的作用是作为基因调节蛋白的结合位点，这些调节蛋白在 DNA 上的存在影响了转录起始的速率。这些序列可以位于启动子相邻位置，或在其上游较远位置，甚至位于内含子中或者基因的下游。DNA 环化据认为使基因调节蛋白能够结合这些位置中任何一个，以便和启动子上组装的蛋白质相互作用。尽管在启动子上装配通用转录因子对于所有聚合酶 II 转录的基因都相同，但对于每个基因而言，基因调节蛋白以及它们结合位点相对于启动子的位置是不同的

本章我们一般用**基因**这个词仅仅指转录成 RNA 的一个 DNA 片段（图 7-41）。然而，基因的经典观点也包括基因调控区。历史上以不同的方法鉴定出基因时就会出现不同的定义。选择性 RNA 剪接的方法使基因的定义更加复杂，这一点我们在第 6 章已简略讨论过，本章后面我们再回来讨论。

虽然许多基因调节蛋白结合增强子序列，并激活基因转录，但是许多其他的基因调节蛋白作为负调节蛋白发挥功能，下面我们会看到。通用转录因子是在 RNA 聚合酶 II 转录的所有基因启动子上组装的丰富蛋白质，与这些种类不多的蛋白质相比，存在数千种不同的基因调节蛋白。例如，在大约 30 000 个人类基因中，估计 5%～10%编码了基因调节蛋白。这些调节蛋白在不同的基因调控区之间不同，每一个通常以少量存在于细胞内，常常不到总蛋白质的 0.01%。它们大多数都使用前面讨论的 DNA 结合基序之一识别特异的 DNA 序列，尽管它们有些并不直接识别 DNA 而是在其他的 DNA 结合蛋白上组装，这些我们将在下面讨论。

基因调节蛋白允许生物的单个基因特异性地打开或关闭。不同的细胞中存在不同的基因调节蛋白选择，因而指导了赋予每个细胞类型独特特征的基因表达模式。真核细胞中几乎每个基因的调控都不相同。考虑到真核生物基因的数量以及它们调控的复杂性，还难以概括出一个适用于所有情况的简单基因调控规则。然而，关于基因调节蛋白一旦结合到 DNA 上的基因调控区就怎样影响转录起始的速率，我们现在就能做一些概括。

真核基因激活蛋白促进 RNA 聚合酶和通用转录因子在转录起始位点上的组装

大多数激活基因转录的基因调节蛋白，也是大部分**基因激活蛋白**，有一个模块化结构，

由至少两种不同的结构域组成，其中一个结构域通常含有前面讨论的识别特异调节 DNA 序列的结构基序之一。在最简单的情况下，另一个结构域（有时称为激活结构域）加快了转录起始速率。利用基因工程技术，构建一个杂合蛋白质，含有一个蛋白质的激活结构域和另一个不同蛋白质的 DNA 结合结构域融合，这样的实验最先揭示了这种模块化结构（图 7-42）。

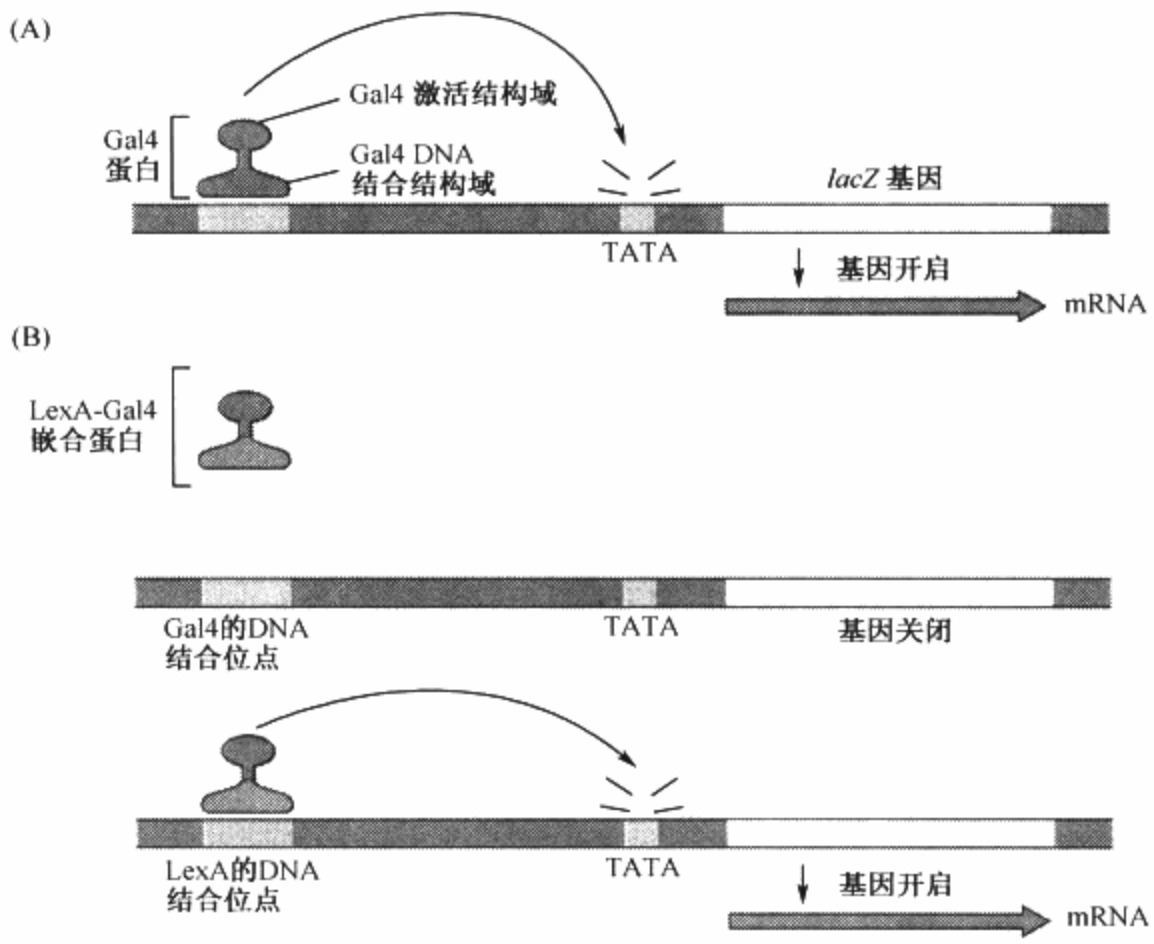


图 7-42　基因激活蛋白的模块化结构

解释酵母基因激活蛋白 Gal4 中存在独立的 DNA 结合结构域和转录激活结构域的实验概图。如果通过基因融合技术将酵母 Gal4 蛋白 C 端部分连接到细菌基因调节蛋白（LexA 蛋白）的 DNA 结合结构域，就可以重建出一个有功能的激活蛋白。当在酵母中合成细菌-酵母杂交蛋白时，它将激活酵母基因的转录，假设已经在基因相邻位置插入了细菌蛋白的特异性 DNA 结合位点。(A) Gal4 蛋白产生的基因转录正常激活 (B) 嵌合基因调节蛋白的活性需要 LexA 蛋白 DNA 结合位点。Gal4 通常负责激活编码将半乳糖转化为葡萄糖的酶的酵母基因转录。在这里所示的实验中，这些基因之一的调控区和大肠杆菌 *lacZ* 基因融合，后者编码 β-半乳糖苷酶（图 7-38）。β-半乳糖苷酶很容易用生物化学方法检测，因此提供了一种方便的途径来监控基因调控区指令的表达水平；因而 *lacZ* 可作为报道基因，因为它“报道”了基因调控区的活性

一旦结合 DNA 后，真核基因活化蛋白怎样提高转录起始速率呢？不久我们将看到，提高转录起始速率以多种机制进行，而且在许多情况下，这些不同的机制在单个启动子上协同发挥作用。但是，不管其精确的生物化学途径是什么，活化蛋白的主要功能就是在启动子上吸引、定位和修饰通用转录因子和 RNA 聚合酶 Ⅱ，这样就能开始转录。它们通过直接作用于转录装置本身和改变启动子附近染色质结构来做到这一点。

首先我们考查活化蛋白直接影响通用转录因子和 RNA 聚合酶 Ⅱ 在启动子上的定位从而帮助促进它们发挥作用的方式。虽然通用转录因子和 RNA 聚合酶 Ⅱ 在体外是逐步地按规定的顺序组装的（图 6-16），可是在活细胞内的确存在一些例子，在此情况下它们有些是作为一个巨大的前组装复合物，有时称作 RNA 聚合酶 Ⅱ 全酶，被带至启动子。除了一些通用转录因子和 RNA 聚合酶外，全酶典型地拥有一个含 20 个亚基的蛋白质复合物，称为中介蛋白，最先从生物化学方面鉴定表明它是刺激转录起始的活化蛋白所必需的。

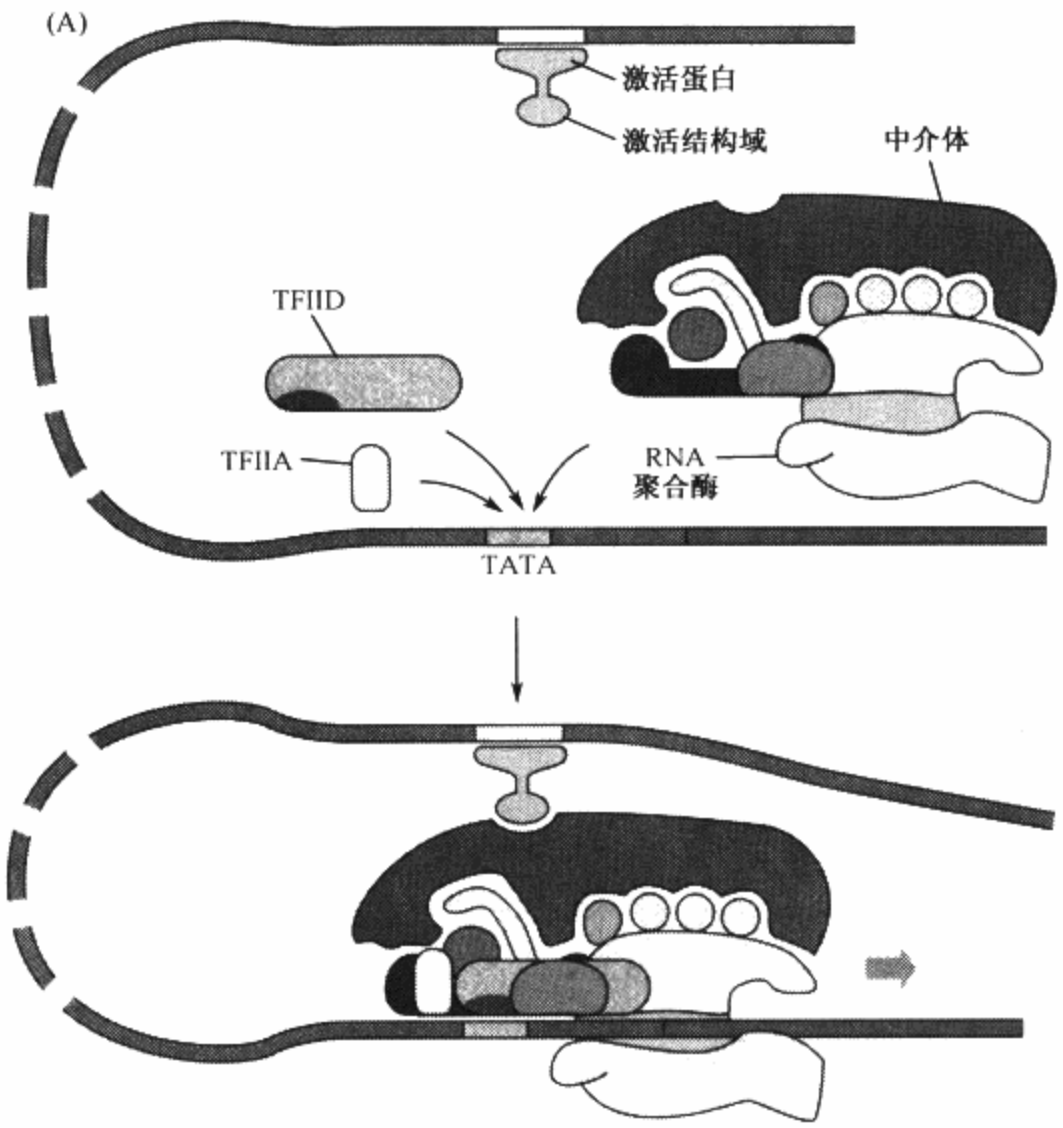


图 7-43　真核生物中通过募集真核 RNA 聚合酶Ⅱ全酶复合体激活转录起始

（A）启动子附近结合的激活蛋白吸引全酶复合体到启动子上来。根据这个模型，全酶（含有 100 多个蛋白质亚基）分别从通用转录因子 TFⅡD 和 TFⅡA 带至启动子。该图以及下图中“断裂”DNA 表示这段 DNA 分子可以非常长，并且长度可以不同。（B）体内实验图，该实验结构支持基因激活蛋白的全酶募集模型。蛋白质的 DNA 结合结构域已直接与中介体的蛋白质组分融合，该中介体是一个含有 20 个亚基的蛋白质复合体，是全酶复合体的一部分，但容易与全酶剩余组分解离。当杂合蛋白的结合位点通过实验方法插入到启动子附近时，就大大增加了转录的起始。该实验中，已删去激活蛋白的“激活结构域”（图 7-42），表明激活结构域一个重要功能只是与 RNA 聚合酶全酶相互作用，因而帮助它在启动子上的组装。利用染色质免疫沉淀法（图 7-32）也能直接证明基因激活蛋白募集转录装置到启动子上的能力。DNA 结合激活蛋白典型地增加转录效率达 1000 倍，这与激活蛋白和全酶之间相对微弱而非特异相互作用是一致的（亲和力改变 1000 倍对应 ΔG 改变约 4kcal/moL，这可以利用只有少量微弱的非共价键来做出解释）

许多活化蛋白与全酶复合物相互作用，因而使其在启动子上的组装在能量上更加有利，该启动子通过 DNA 连接到活化蛋白结合的位点上[图7-43(A)]。在这种意义上讲，真核活化蛋白与细菌活化蛋白在帮助吸引并定位 RNA 聚合酶到 DNA 的特异位点上是相似的（图 7-36）。有一种支持活化蛋白吸引全酶复合物到启动子这种观点的实验，创造出了一种“活化蛋白旁路”[图7-43(B)]。这里，利用实验方法将一个序列特异的 DNA 结合结构域直接与中介体的一个成分融合；当将这个缺少激活结构域的杂合蛋白结合的 DNA 序列放到一个启动子附近时，它就会强有力地刺激转录起始。

虽然全酶复合物在启动子上的募集为基因激活过程提供了一种在理论上简单的机制，但是活化蛋白对全酶复合物的影响可能更加复杂。例如，在某些启动子上可能发生通用转录因子（图 6-16）的逐步组装。在另一些启动子上，通用转录因子一旦作为全酶的一部分带到 DNA 上，就有可能需要它们的重排。此外，大部分全酶复合物形式缺少某些通用转录因子（最显著的是 TFⅡD 和 TFⅡA），这些必须分别在启动子上组装[图7-43(A)]。从理论上讲，这些组装过程中任何一个都可能是通向转录起始途径中的慢步骤，而活化蛋白能够促进这些过程的完成。实际上，已经表明许多活化蛋白与一个或更多的通用转录因子相互作用，还表明有许多活化蛋白直接促进了它们在启动子上的组装（图 7-44）。

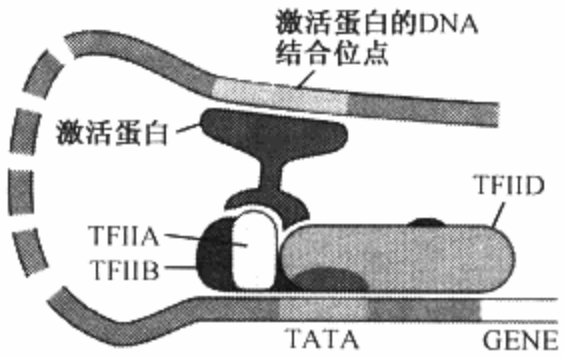


图 7-44　解释某些真核转录激活蛋白作用的模型

结合在启动子附近的基因激活蛋白，促进某些通用转录因子的装配。虽然有些激活蛋白可能只在转录起始途径特定的步骤中发挥作用，但许多似乎能在多个步骤中发挥功能

真核基因激活蛋白修饰局部染色质结构

基因激活除了直接发挥作用，将 RNA 聚合酶全酶和通用转录因子在 DNA 上组装之外，也通过改变基因的调节序列和启动子染色质结构促进转录起始。我们在第 4 章已看到过，局部改变染色质结构的两种最重要的方式是通过共价组蛋白修饰和核小体重建（图 4-34 和图 4-35）。许多基因激活蛋白通过结合并因而募集组蛋白乙酰转移酶（HAT)(通常称作组蛋白乙酰酶）和依赖于 ATP 的染色质重建复合体（图 7-45），利用了这两种机制来作用于附近的染色质。一般来讲，随之发生的染色质结构局部改变使得接触其中的 DNA 更加容易。这个可接触性促进了通用转录因子和 RNA 聚合酶全酶在启动子上的组装，它也允许其他的基因调节蛋白结合到基因调节区[图7-46(A)]。

通用转录因子似乎不能在已包装到常规核小体中的 DNA 上组装。事实上，这种包装可能部分已经演变过，以确保不会发生渗漏的或基础的转录起始（启动子上游没有结合基因活化蛋白时发生的起始）。除了使 DNA 更容易广泛地接触之外，局部组蛋白乙酰化在促进转录起始过程中发挥更加特殊的作用。第 4 章已经讨论过（图 4-35），某些组蛋白乙酰化模式与转录活跃的染色质相关联，并且基因活化蛋白通过募集组蛋白乙酰化酶产生了这些乙酰化模式。其中一种这样的模式[图7-46(B)]直接由通用转录因子 TFⅡD的一个亚基识别，这种识别显然有助于该因子与包装成染色质的 DNA 进行组装。因此通过组蛋白乙酰酶的作用，基因活化蛋白能间接帮助通用转录因子在启动子上的组装，因而刺激转录起始。

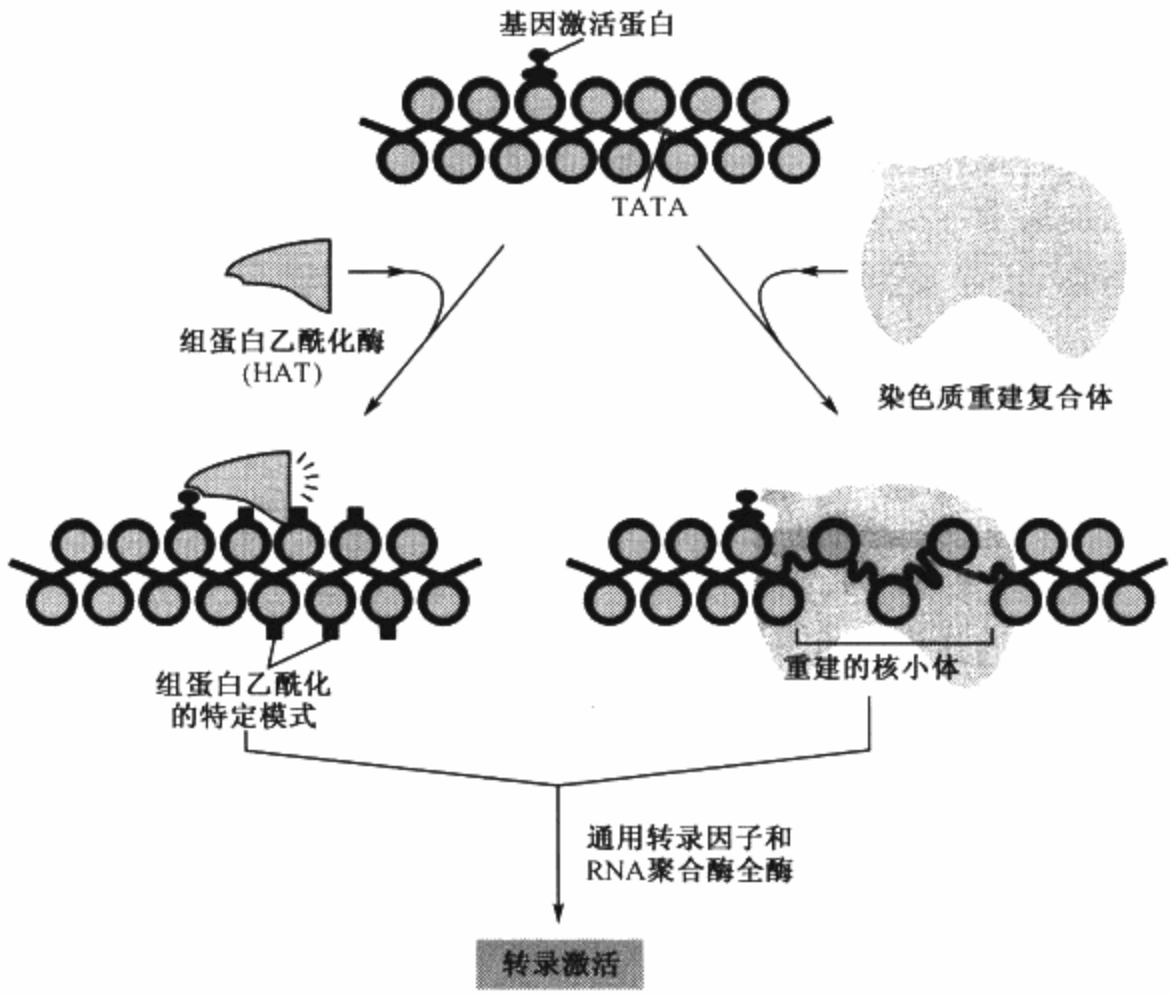


图 7-45　真核基因激活蛋白引导的染色质结构局部改变

组蛋白乙酰化和核小体重建通常使得包装在染色质中的 DNA 更容易被细胞中的其他蛋白质所接触，包括转录起始所需的蛋白质。此外，组蛋白修饰的特定模式直接帮助通用转录因子在启动子上组装（图 7-46）。转录起始和紧密的染色质结构的形成可以看作是竞争性的生物化学组装反应。酶的增加，即使是短暂性地增加染色质中 DNA 可触及性的酶将倾向作用于转录起始（图 4-34）

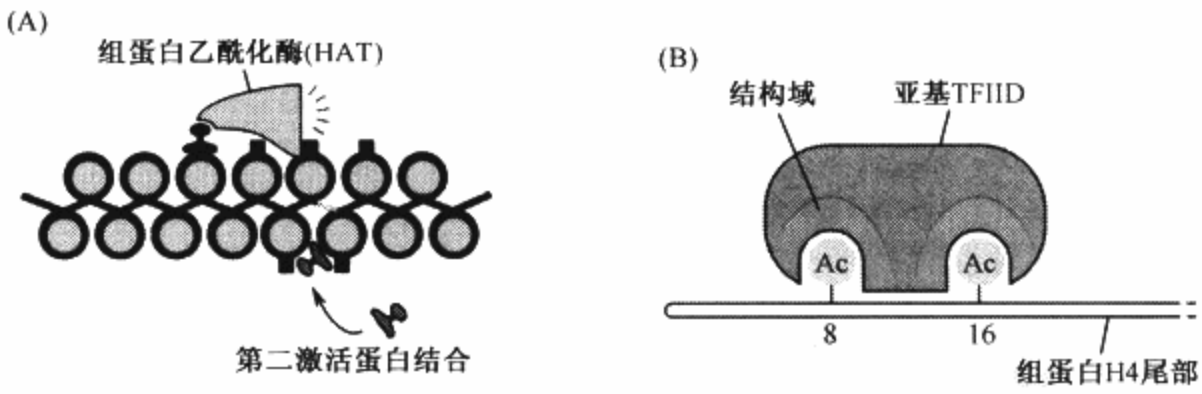


图 7-46　局部组蛋白乙酰化促进转录起始的两种特殊方式

（A）有些基因调节蛋白能直接结合包装在未修饰染色质中的 DNA。通过吸引组蛋白乙酰化酶（和核小体重建复合体），这些“先驱”激活蛋白能促进不能结合未修饰染色质的其他激活蛋白结合到 DNA 上。这些其他蛋白质继而能对染色质进行其他的修饰或者直接作用于如图 7-43 和图 7-44 中所示的转录装置。（B）通用转录因子TFⅡD的亚基含有两个称作同源调节域的含有 120 个氨基酸的蛋白质结构域。每个同源调节域形成一个乙酰化赖氨酸侧链的结合口袋（图中命名为 Ac）；在 TFⅡD 中这两个口袋相隔 25Å，这是识别组蛋白 H4 N 端尾上 6 或 7 个氨基酸分隔开的一对乙酰化赖氨酸的最佳间距。除了所显示的乙酰化模式外，TFⅡD 的这个亚基也识别在 5 位和 12 位乙酰化的组蛋白 H4 尾部以及完全乙酰化的尾部。它对未乙酰化的 H4 尾部没有明显的亲和力，对只在一个赖氨酸上乙酰化的 H4 尾部亲和力很低。如图 4-35 中所示，某些组蛋白 H4 乙酰化的模式，包括由 TFⅡD 识别的模式，与染色质转录活性区域相关联

基因激活蛋白协同地发挥作用

我们已经看到真核基因激活蛋白能影响转录起始过程中多个不同步骤，当不同的激活蛋白一起发挥作用时这种特性就产生重要的结果。一般来讲，多个因子共同作用来增加反应速率的地方，联合效应通常不仅仅是每个因子单独增加量的总和，而是它们的乘积。例如，如果因子 A 降低某个反应一定数量的自由能障碍，因而使反应加速了 100 倍，因子 B 通过作用于该反应的另一方面，产生同样的结果，那么 A 和 B 并联作用将双倍地降低自由能障碍，并提高反应速度达 10 000 倍。如果 A 和 B 各自通过帮助募集必需的蛋白质到反应位点也会以类似的乘积效应增加反应速度。因此，基因激活蛋白常常显示出一种称作转录协同的特征，这里由多种激活蛋白一起作用产生的转录速率比任何激活蛋白独自作用产生的转录速率要大得多（图 7-47）。在结合基因上游的不同基因激活蛋白之间以及相同激活蛋白的多个 DNA 结合分子之间都能观察到转录协同。因此不难看到各自结合不同调节 DNA 序列的多个基因调节蛋白是怎样调控真核基因最终转录速率的。

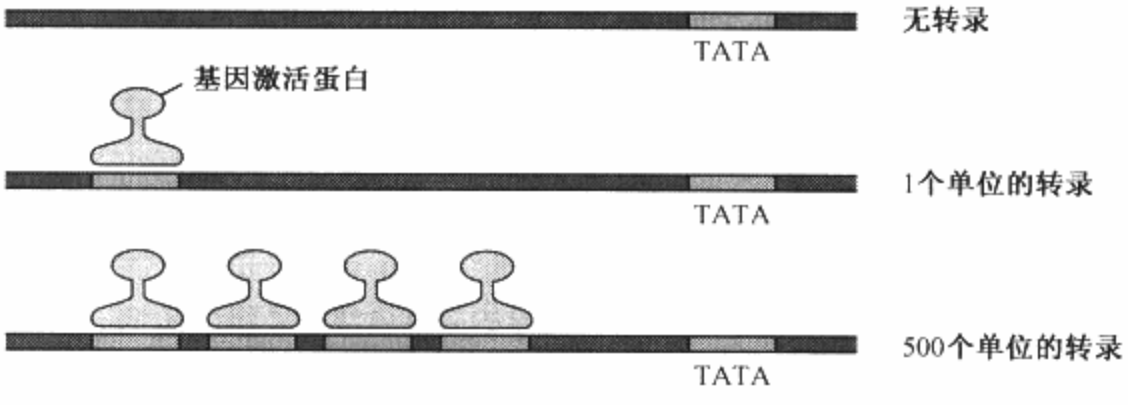


图 7-47　转录协同

该实验中，比较了 3 种通过实验手段构建的调控区在真核细胞中产生的转录速率。当多个基因激活蛋白分子结合到启动子上游时可以观察到转录协同效应，它比激活蛋白的加和效应更大。来自同一生物的不同基因激活蛋白之间，甚至来自广泛的不同真核生物物种的激活蛋白之间，如果通过实验手段引入到同一细胞内，通常也可以观察到协同作用。最后一种情况下观察到的协同作用反映了转录装置的高度保守性

既然基因调节蛋白能够影响通向转录活化途径中的许多不同步骤，就值得考虑一下是否这些步骤总是按规定的顺序进行的。例如，染色质重建必须在组蛋白乙酰化之前进行吗？或者反之亦然。相对于染色质修饰步骤而言全酶复合物的募集是何时发生的？这些问题的答案对于不同的基因似乎是不同的，甚至是对同一基因在不同条件下也是不同的（图 7-48）。无论其精确的机制以及它们进行的顺序是什么，一种基因调节蛋白为了影响它的目标启动子的转录，必须直接或间接地结合到 DNA 上，并且一个基因最终转录速率依赖于其转录起始位点上下游结合的多种调节蛋白。

真核基因阻遏蛋白能以不同的方式抑制转录

与细菌相似，除了激活蛋白以外，真核生物也利用**基因阻遏蛋白**调节它们的基因转录。然而，由于真核生物和细菌转录起始的方式不同，真核阻遏蛋白可能的作用机制要多得多。例如，我们在第 4 章看到，真核生物染色体整个区域能包装成异染色质，它是一种通常能抵制转录的染色质形式。本章后面我们将再回来讨论真核生物染色体的这个特征。除了关闭大段染色质区域的分子以外，真核细胞也拥有只在局部发挥作用抑制附近基因转录的基因调节蛋白。与细菌阻遏蛋白不同，真核基因阻遇蛋白大部分都不直接与 RNA 聚合酶竞争接触

DNA；而是以多种其他的机制工作，其中一些在图 7-49 中加以说明。与基因激活蛋白一样，许多真核生物阻遏蛋白不只以一种机制作用，因而确保了强有力而且有效的阻遏效果。

真核基因调节蛋白常常在 DNA 上装配成复合体

到目前我们一直在讨论真核基因调节蛋白，似乎它们是以单独的多肽链起作用的。事实上，大多数是作为几个多肽链（有时是许多）组成的复合体的一部分，每个都有不同的功能。这些复合体通常只在合适的 DNA 序列存在时才组装起来。在有些研究得比较清楚的例子中，例如，两个彼此亲和力微弱的基因调节蛋白协同起来结合一段 DNA 序列，任何一个蛋白质单独都不具备有效结合 DNA 位点足够的亲和力。一旦结合上 DNA，这个蛋白质二聚体形成不同的表面，由第三个携带刺激转录的激活结构域的蛋白质进行识别（图 7-50）。这个例子说明了一个重要的普遍观点：太微弱而不能使蛋白质在溶液中组装的蛋白质-蛋白质相互作用能使蛋白质在 DNA 上组装；在这种方式中 DNA 作为“结晶”位点或蛋白质复合体组装的种子起作用。

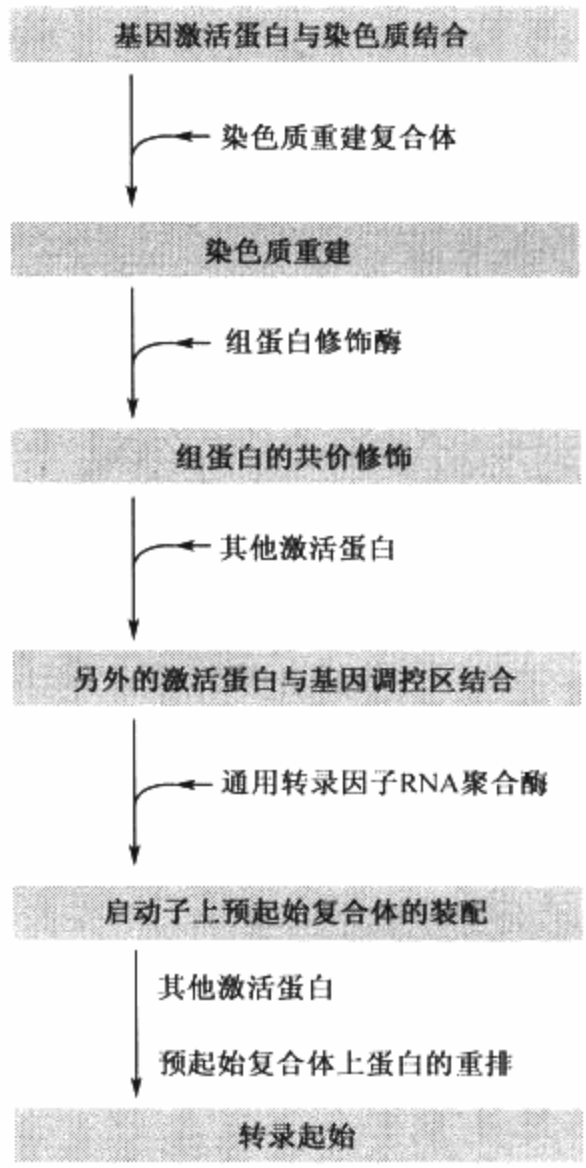


图 7-48　在特定启动子上引发转录起始的一系列事件的先后顺序

图中所示的经过详细研究的例子来自于酿酒酵母的一个启动子。染色质重建复合体和组蛋白乙酰酶在它们连续作用以后显然从 DNA 上解离。在通向转录起始的途径中，反应步骤的顺序对于不同启动子似乎不同。例如，在来自人类的一个经过详细研究的例子中，组蛋白乙酰酶最先发挥功能，然后募集 RNA 聚合酶，再募集染色质重建复合体

单个基因调节蛋白常常能参与形成不只一种调节复合体。例如，一个蛋白质可能在一种情况下作为激活转录的复合体组分，在另一种情况下则作为阻遏转录的复合体组分（图 7-50）。因此单个真核基因调节蛋白并不一定是用作激活蛋白或者阻遏蛋白；反之，它们的作用是作为一个调节基本单位用于形成复合体，其功能由全部单个成分最终的组装决定。继而这种最终的组装再决定于调控区域 DNA 序列的排列和细胞内基因调节蛋白存在于哪个上面。

自身不结合 DNA 但与 DNA 结合基因调节蛋白组装的基因调节蛋白通常称作**共激活蛋白**或**共阻遏蛋白**，由它们对转录起始的效应决定。如图 7-50 所示，相同的共激活蛋白或共阻遏蛋白能够在不同的 DNA 结合蛋白上组装。共激活蛋白或共阻遏蛋白典型地具有多种功能：它们能与染色质重建复合体、组蛋白修饰酶、RNA 聚合酶全酶和多种通用转录因子相互作用。

在有些情况下，一个调节蛋白直接结合的精确 DNA 序列能影响这个蛋白质的构象，因而会影响它随后的转录活性。例如，一种类固醇激素受体当与某种类型 DNA 序列结合后，它与共阻遏蛋白相互作用并最终关闭转录。当与一种略微不同的 DNA 序列结合时，它呈现一种不同的构象并与共激活蛋白相互作用，因而刺激了转录。

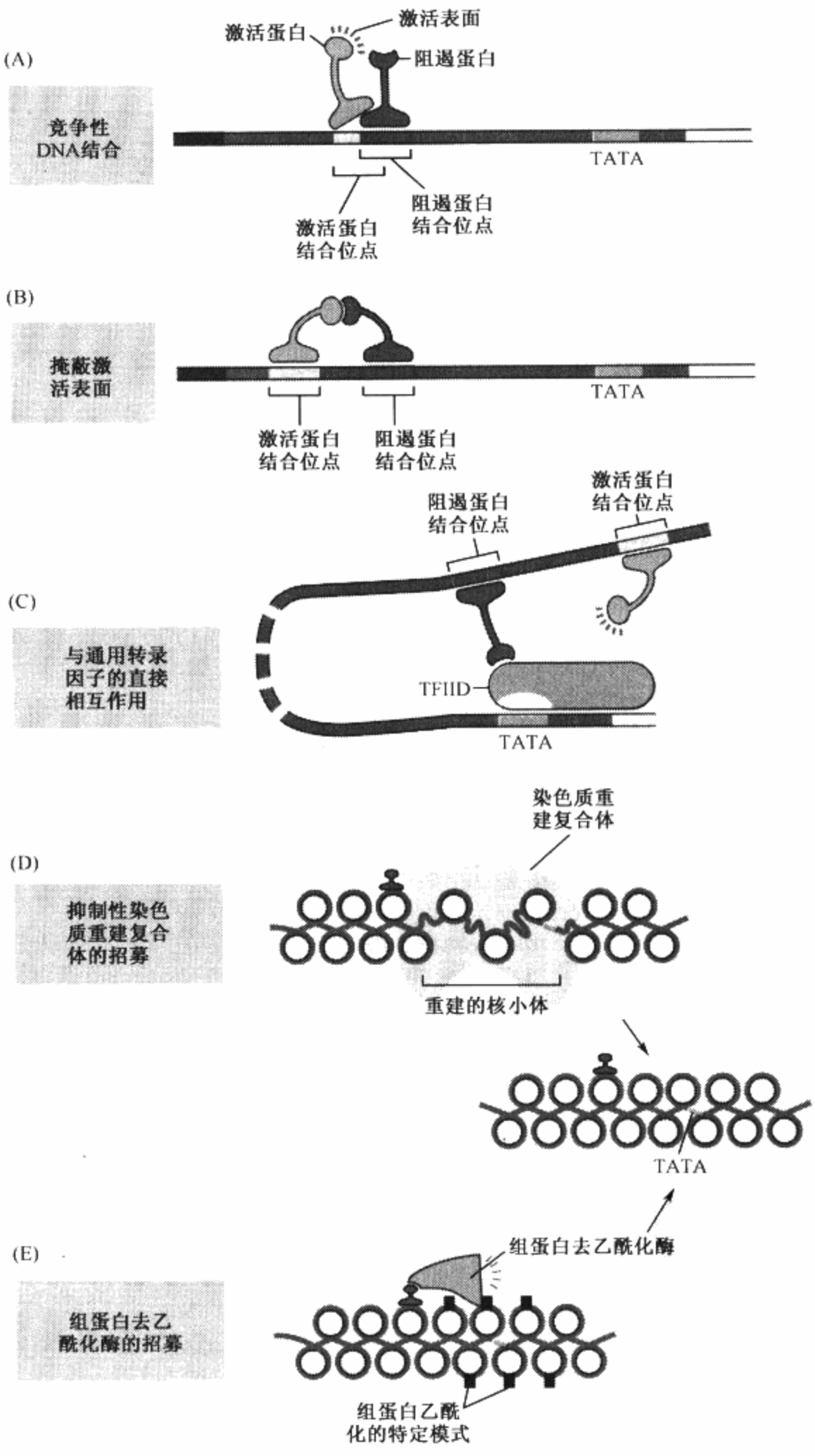


图 7-49　真核基因阻遏蛋白发挥作用的 5 种方式

(A) 基因激活蛋白与基因阻遏蛋白竞争结合同一调节 DNA 序列。(B) 两种蛋白质都能结合 DNA，但阻遏蛋白结合到激活蛋白的激活结构域，因而阻止它行使激活功能。在这种策略的一种变通形式中，阻遏蛋白牢固结合激活蛋白，而不必直接结合 DNA。(C) 阻遏蛋白与通用转录因子早期组装复合物相互作用，阻断其进一步组装。有些阻遏蛋白也在转录起始后期发挥作用。例如，通过阻止 RNA 聚合酶从通用转录因子上释放。(D) 阻遏蛋白募集染色质重建复合体，它使启动子区域核小体状态返回到转录前形式，有些类型的重建复合体似乎专门用于恢复启动子受阻的核小体状态，而另一些（如那些激活蛋白募集的重建复合体）使得包装在核小体中的 DNA 更易于接触（图 4-34）。然而从理论上讲相同的重建复合体能用于或者激活或者阻遏转录：根据细胞核内其他蛋白质的浓度，重建状态或受阻遏状态都能稳定下来。根据这个观点，重建复合体只是使染色质结构能够发生改变。(E) 阻遏蛋白吸引组蛋白脱乙酰化酶到启动子上。局部的组蛋白脱乙酰化减少 TF Ⅱ D 对启动子的亲和力（图 7-46），并降低受影响的染色质中 DNA 可接触性。图 7-26 说明了第 6 种负调控机制——转录激活蛋白通过异源二聚体化失活。为了简单起见，(A)～(C)中删去了核小体，(D)和(E)所用的比例尺比(A)～(C)的要小

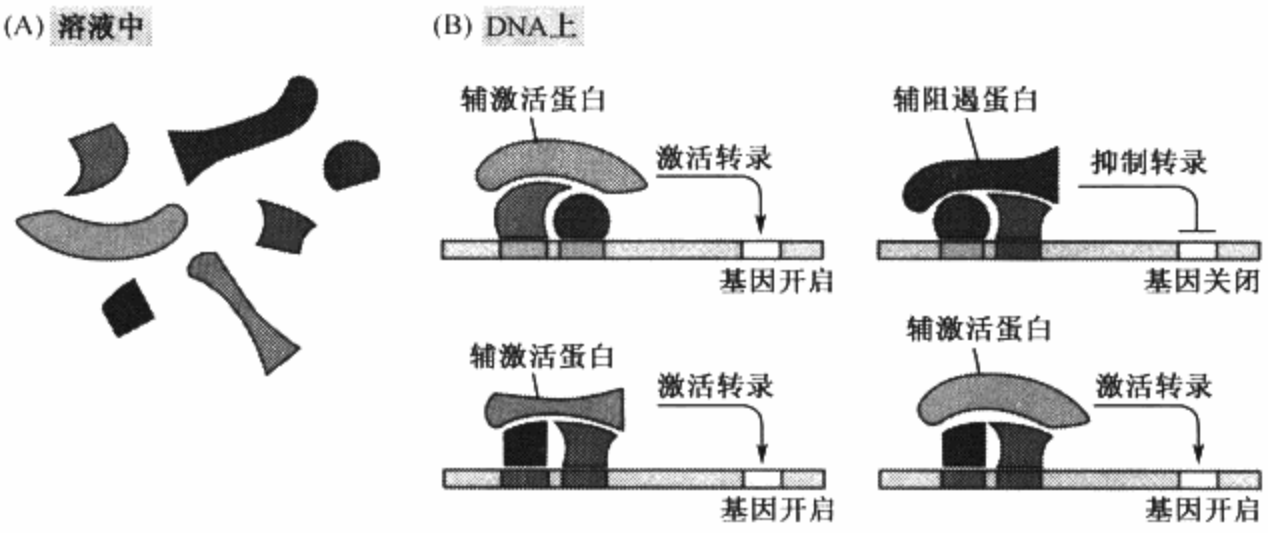


图 7-50　真核基因调节蛋白常常在 DNA 上组装成复合体

（A）中显示了 7 种基因调节蛋白，它们形成的复合体的特性和功能决定于提供它们组装的特异 DNA 序列。（B）中有些组装的复合体激活基因转录，而另一些阻遏基因转录。注意红色蛋白质为激活复合体和阻遏复合体所共有

通常一组调节蛋白在 DNA 上的组装最终受少量较短的核苷酸序列片段引导（图 7-50）。然而有些情况下，形成了一种更精巧的蛋白质-DNA 结构，称为增强体（图 7-51）。增强体的一个检验印记就是构筑蛋白的参与，它们以一定的角度弯曲 DNA，因而促进了其他增强子蛋白质的组装。由于增强体的形成要求存在许多基因调节蛋白，所以它提供了一种简单的方式来确保只有当细胞内存在这些蛋白质的正确组合时，才会表达某个基因。我们前面已经看到溶液中基因调节异源二聚体的形成怎样提供一种基因表达组合调控的机制。在 DNA 上组装更大的基因调节蛋白复合物提供了另一种重要的组合调控的机制，从而为细胞提供了更多的机会。

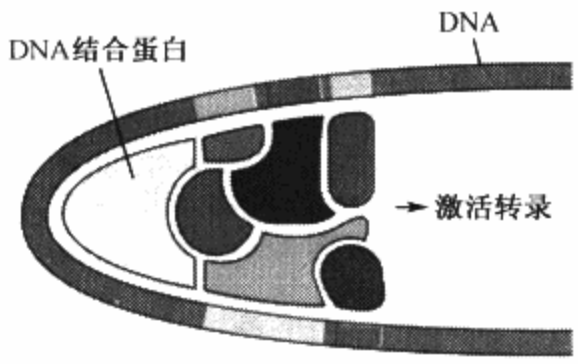


图 7-51　增强体的简图

用黄色表示的蛋白质称为构筑蛋白，因为它的主要作用是将 DNA 弯曲，以便其他组分能协同组装。该增强体的蛋白质表面与在一个启动子附近激活转录的共激活物相互作用。这里描绘的增强体基于在 T 细胞受体一个亚基的编码基因调控区中发现的增强体（讨论见第 24 章）。增强体完整的一套蛋白质组分只存在于正在发育的免疫系统某些细胞中，这些细胞最终形成成熟的 T 细胞

调节果蝇发育的复杂基因开关由较小的元件构造而成

考虑到基因调节蛋白可以定位到长 DNA 片段上的多个位点，这些蛋白质在每个位点上都装配成复合体，并且该复合体能影响染色质的结构，以及通用转录装置的募集和在启动子上的组装，这样看起来似乎调控真核基因转录的装置非常复杂。

这种复杂多组分基因开关的一个显著例子就是果蝇偶跳基因（*eve*）的转录调控，这种基

因的表达在果蝇胚胎发育过程中起着重要作用。如果这个基因突变失活，那么胚胎的许多部位都不能形成，在发育早期就会死亡。第 21 章将会讨论，在 *eve* 进行表达的发育最早阶段，此时胚胎是单个巨大的细胞，共同的细胞质内含有多个细胞核。然而这种细胞核并不一致：它含有基因调节蛋白的混合物，这些蛋白质沿胚胎长轴方向不均匀分布，因此提供了一种定位信息，区分胚胎的不同部分（图 7-52)(第 21 章讨论了最初确立这些差别的方式)。虽然最初细胞核是相同的，但是它们迅速地开始表达不同的基因，因为它们接触了不同的基因调节蛋白。例如，发育胚胎前端附近的细胞核接触的一组基因调节蛋白与影响胚胎后端细胞核的一组调节蛋白不同。

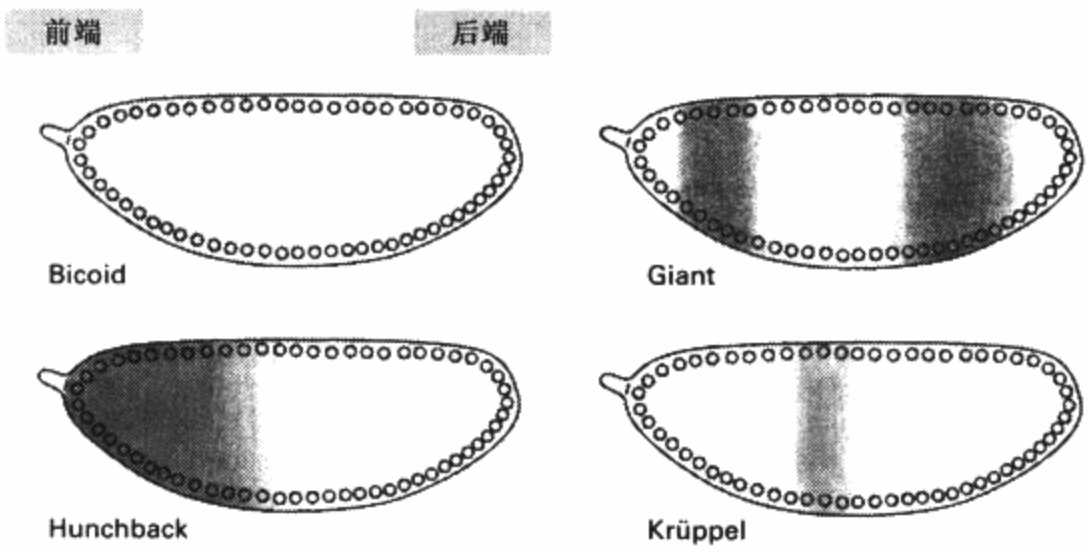


图 7-52　早期果蝇胚胎中 4 个基因调节蛋白不一致的分布

在此阶段胚胎是一个合胞体，多个细胞核存在于共同的细胞质中。虽然未在这些图中加以说明，但所有这些蛋白质都集中在细胞核内

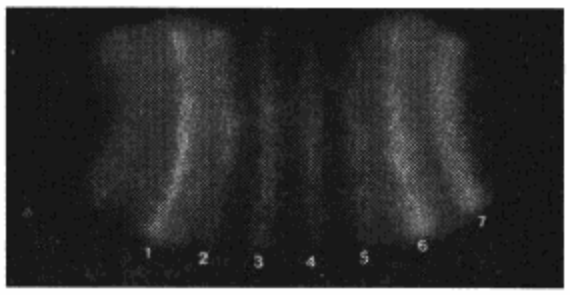


图 7-53　果蝇发育胚胎中偶跳基因 (*eve*) 编码的蛋白质的 7 条带

受精后 2.5h，将识别 Eve 蛋白（绿色）的抗体和识别 Giant 蛋白（红色）的抗体注入卵中并对其染色。在 Eve 和 Giant 蛋白都存在的区域，染色呈黄色。在发育的这一阶段，卵含有将近 4000 个细胞核。Eve 蛋白和 Giant 蛋白都位于细胞核，Eve 带大约有 4 个细胞核宽。图 7-52 中也显示了 Giant 蛋白的染色模式（经 Michael Levine 许可）

eve 基因的调节 DNA 序列能够阅读沿胚胎长轴方向的每个位置的基因调节蛋白浓度，并以这样一种方式解释该信息：*eve* 基因在 7 个带区中表达，每个带区最初有 5 或 6 个细胞核宽，并沿胚胎前后轴精确地定位（图 7-53）。这种信息加工的卓越绝技是如何进行的呢？虽然还不了解全部的分子细节，但是从对 *eve* 和其他受类似调控的果蝇基因的研究中得到了一些一般性规则。

eve 基因的调控区很大（将近 20 000 个核苷酸对）。它由一系列比较简单的元件组成，每个元件含有多种调节序列并负责指令在胚胎上 *eve* 表达的特定条带。将特定的一个调节模块（如指令条带 2 的模块）从其在 *eve* 基因上游所处正常位置上移走，并放到一个报道基因前面（图 7-42），再重新引入到果蝇基因组[图7-54(A)]。对携带这种基因构造的果蝇繁衍的发育胚胎进行检测，发现报道基因就在条带 2 的位置表达（图 7-54）。类似的实验解释了其他调节模块的存在，它们各自指令其他 6 个条带中的一条或者该基因在发育后期展示的表达模式的一部分。

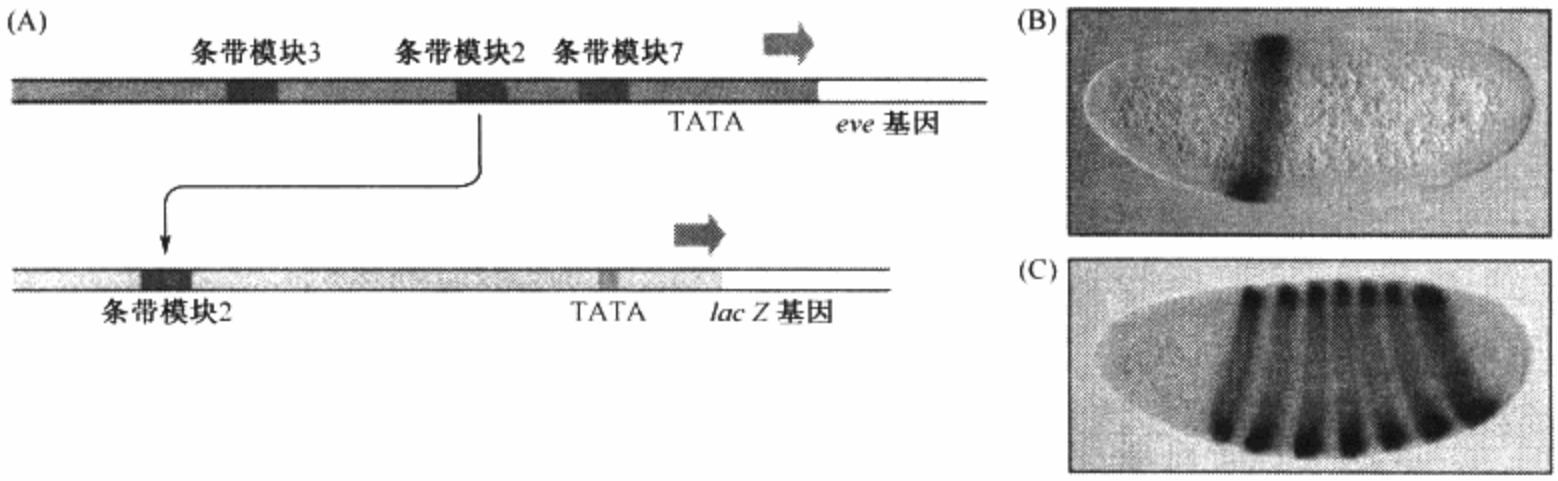


图 7-54　证明 *eve* 基因调控区模块式构造的实验

（A）将 *eve* 调控区一段长 480 个核苷酸对的片段移走，并插入到指导 β-半乳糖苷酶合成（大肠杆菌 LacZ 基因产物）的试验启动子上游。(B) 当这个人造产物重新引入果蝇胚胎基因组后，该胚胎在 7 条 *eve* 带（C）的第二条中精确地表达 β-半乳糖苷酶（可用组织化学染色方法观察到）[(B)和(C)经 Stephen Small and Michael Levine 许可]

果蝇 *eve* 基因受到组合调控的调节

对条带 2 调节模块的详细研究帮助我们更深刻地认识了它是怎样阅读和解释位置信息的。它分别含有两个激活 *eve* 转录的基因调节蛋白（Bicoid 和 Hunckback）的识别序列和两个阻遏它转录的阻遏蛋白（Kruppel 和 Giant）的识别序列（图 7-55）。（果蝇的基因调节蛋白常常拥有丰富多样的名字，它们反映了编码该蛋白质的基因由于突变失活后产生的表型。）这 4 种蛋白质的相对浓度决定在条带 2 形成的蛋白质复合物是否打开 *eve* 基因转录。图 7-56 显示了 4 种基因调节蛋白区域的分布，该区域是条带 2 在果蝇胚胎上形成的区域。虽然不知道精确

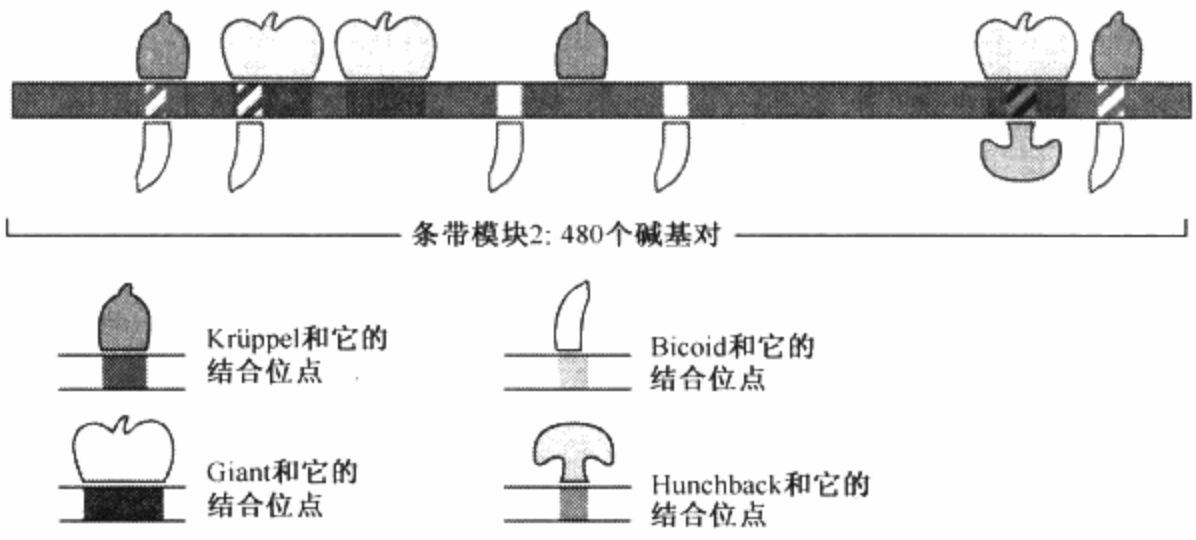


图 7-55　条带 2 中 *eve* 基因的近距离图

在前面的图中鉴定的 *eve* 基因调控区片段含有调控序列，每个都结合 4 个基因调节蛋白中的某一个。从遗传学实验知道了这 4 个基因调节蛋白负责 *eve* 在条带 2 中的正确表达。例如，缺乏两种基因激活蛋白 Bicoid 和 Hunchback 的果蝇不能在条带 2 中有效表达 *eve*。缺少 Giant 和 Kruppel 这两种基因阻遏蛋白中任一个的果蝇中，条带 2 就会扩增并覆盖胚胎异常宽的区域。这些基因调节蛋白的 DNA 结合位点的确定是通过克隆编码这些蛋白质的基因，在大肠杆菌中过量表达它们，然后纯化，并按第 8 章描述的那样进行 DNA 足迹实验。上面的图表示在有些情况下，基因调控蛋白的结合位点发生重叠，这些蛋白质就能竞争与 DNA 结合。例如，在最右边的位点上 Kruppel 的结合和 Bicoid 的结合被认为是互相排斥的

的细节，但似乎有可能两个阻遏蛋白中任何一个结合上 DNA 以后都将关闭条带 2 模块，而 Bicoid 和 Hunchback 则必须结合以获得最大活性。因而这个简单的调节单位将 4 种位置信号组合起来，以便只在某些 Bicoid 和 Hunchback 水平较高而缺少 Kruppel 和 Giant 的核中打开条带模块 2（因而打开 *eve* 基因的表达）。激活蛋白和阻遏蛋白的这种组合只在早期胚胎的一个区域发生；因此在任何其他的地方，条带 2 都是沉默的。

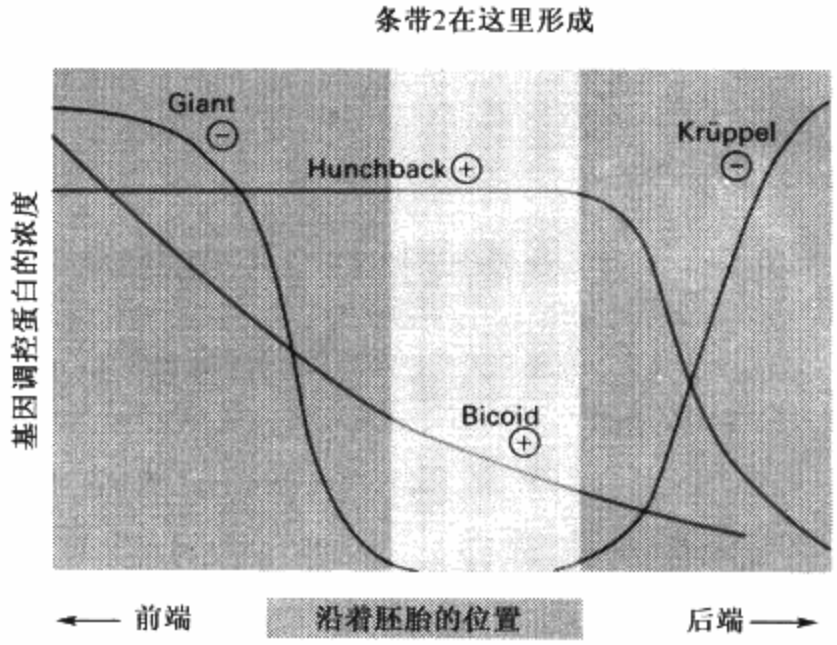


图 7-56　负责确保 *eve* 在条带 2 中表达的基因调节蛋白的分布

这些蛋白质的分布通过用这 4 种蛋白质各自的抗体将发育的果蝇胚胎染色后观察到（图 7-52 和图 7-53）。条带 2 中的 *eve* 只在存在 2 个激活蛋白（Bicoid 和 Hunchback）而缺乏 2 个阻遏蛋白（Giant 和 Kruppel）的位置表达。在缺少 Kruppel 的果蝇胚胎中，条带 2 只在后端扩增。同样，如果条带 2 模块中 Kruppel 的 DNA 结合位点（图 7-55）因突变失活并且这个调控区重新引入到基因组中，那么条带 2 也在后端扩增。*eve* 基因自身编码一种基因调控，这种蛋白质表达模式一旦在 7 个条带中建立，就调节其他果蝇基因的表达。随着发育向前进行，胚胎分为越来越精细的区域，它们最终产生成年果蝇不同的身体部位，这些内容将在第 21 章中讨论。这个来自果蝇胚胎的例子不同寻常，因为细胞核直接接触基因调节蛋白浓缩形式中的定位信号。在大多数其他生物胚胎中，单个细胞核位于单独的细胞中，细胞外的定位信号必须穿过细胞质膜，或者更常见的是，为了影响基因组而在细胞质中产生信号

我们已经讨论了两种基因表达组合调控的机制，基因调节蛋白在溶液中异源二聚化（图 7-22）和基因调节蛋白的组合在 DNA 上装配成小复合体（图 7-50）。可能两种机制都参与了 *eve* 表达的复杂调节。此外，刚才描述的条带 2 的调节说明了第三种组合调控机制。由于 *eve* 条带 2 模块中的单个调节序列沿 DNA 排列成行，所以许多套基因调节蛋白能自发结合并影响一个基因的启动子。启动子整合了所有结合蛋白提供的转录信号（图 7-57）。

eve 表达的调节是组合调控中一个令人印象深刻的例子。基因调节蛋白的 7 种组合（每一条带一种组合）激活了 *eve* 表达，而许多其他组合（在胚胎条带间区域发现的所有组合）使条带元件保持沉默。据认为其他条带调节模块沿着与针对条带 2 描述的类似直线进行构建，这样安排是为了用于阅读基因调节蛋白的其他组合提供的位置信息。整个基因调控区域，沿 DNA 排列长度超过 20 000 个核苷酸对，结合至少 20 种不同的蛋白质。因此一系列较小的模块构建成一个巨大而又复杂的调控区域，每一个小模块又由特定基因调节蛋白识别的短 DNA 序列以独特的方式排列组成。虽然还不了解细节，据认为这种基因调节蛋白运用了前面针对活化蛋白和阻遏蛋白所描述的多种机制。通过这种方式，单个基因能应答大量输入的组合信号。

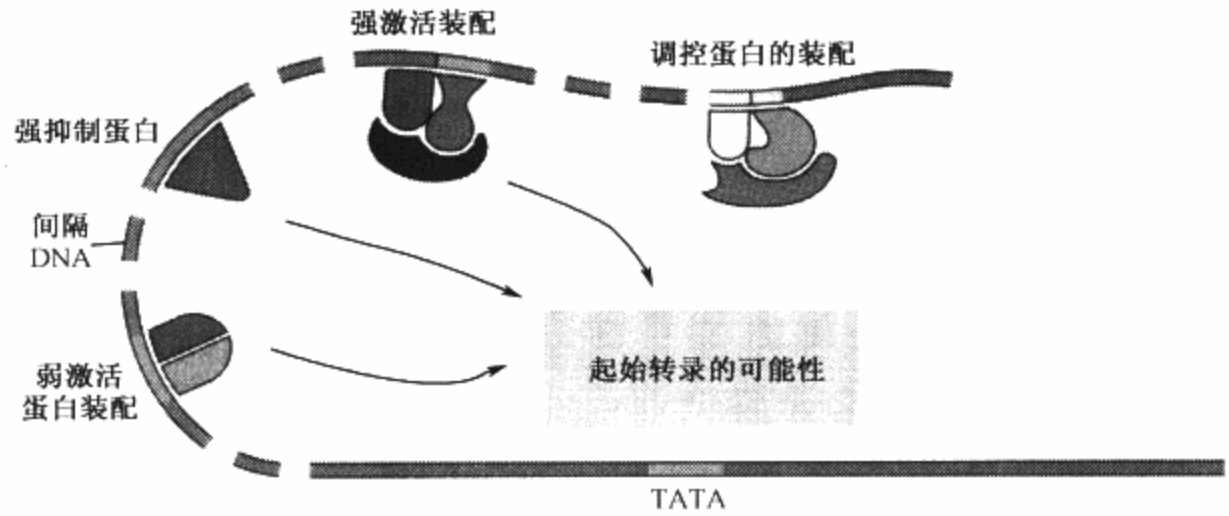


图 7-57　启动子上的整合

多组基因调节蛋白能在启动子上一起发挥作用来影响转录起始，正如在图 7-55 中说明的那样。目前还没有详细地了解是如何实现多种输入信号的整合的，但可能基因最后的转录活性来自于激活蛋白和阻遏蛋白，通过图 7-43～图 7-46 和图 7-49 中总结的机制竞争发挥作用

哺乳动物复杂的基因调控区域也由简单的调节模块构成

据估计哺乳动物基因组中 5%～10%的编码能力都用来合成充当基因转录调节子的蛋白质。如此巨大的基因数目反映了管理哺乳动物基因表达的极为复杂的调控网络。每个基因都受到一组基因调节蛋白调控；这些蛋白质中的每一种都是继而由另一组蛋白质调控的基因产物，依此类推。而且调节蛋白分子自身受到来自细胞外的信号的影响，这些信号使它们以多种不同的方式活化或失活（图 7-58）。因此，细胞中的基因表达模式可以看成是细胞内基因调控网络应答细胞环境的信息所进行的复杂分子计算的结果。我们将在第 21 章针对多细胞发育进一步讨论这一点，但是即使是在调节单个基因活性的单个基因开关水平上，其复杂性也是显著的。例如，找到一个控制区长达 50 000 个核苷酸对的哺乳动物基因并不奇怪，在其中许多模块上有大段间隔 DNA 散布排列，每个模块都含有许多结合基因调节蛋白的调节序列。

了解得最清楚的哺乳动物调节区域的例子在人类 β-珠蛋白基因中发现，这个基因专门在红细胞中且在其特定的发育时间表达。基因调节蛋白的复杂排布调控着基因表达，其中一些作为激活蛋白，而另一些作为阻遏蛋白（图 7-59）。人们认为许多这些基因调节蛋白的浓度（或者活性）在发育期间发生变化，只有全部蛋白质的某个特定组合能引发该基因的转录。人类 β-珠蛋白基因是珠蛋白基因簇的一部分[图7-60(A)]。串联着的这 5 个基因只在类红细胞，即红细胞系的细胞中转录。此外，每个基因在发育的不同时期[图7-60(B)]和不同的器官中打开：ε-珠蛋白基因在胚卵黄囊中表达，γ-珠蛋白在卵黄囊和胎儿肝中表达，而 δ-珠蛋白和 β-珠蛋白主要在成人骨髓中表达。每个珠蛋白基因拥有自己的一套调节蛋白，它们对于在合适的时间和组织内打开该基因来说是必需的。除了每个珠蛋白基因的单独调节以外，似乎整串基因也受一段共有调控区域的调节，称其为基因座控制区（LCR）。LCR 位于远离基因簇的上游位置[图7-60(A)]，我们将接着讨论它的功能。

在珠蛋白基因不表达的细胞中（如脑细胞和皮肤细胞），整个基因簇看起来紧密地包装到染色质中。相比之下，在红细胞中整个基因簇仍然折叠成核小体，但是染色质较高级的包装已经解压缩了。这个变化甚至在单个珠蛋白基因转录之前就会发生，表明存在两步调节。首先，整个珠蛋白基因座的染色质解压缩，这可能使其他的基因调节蛋白能够接触 DNA。第二

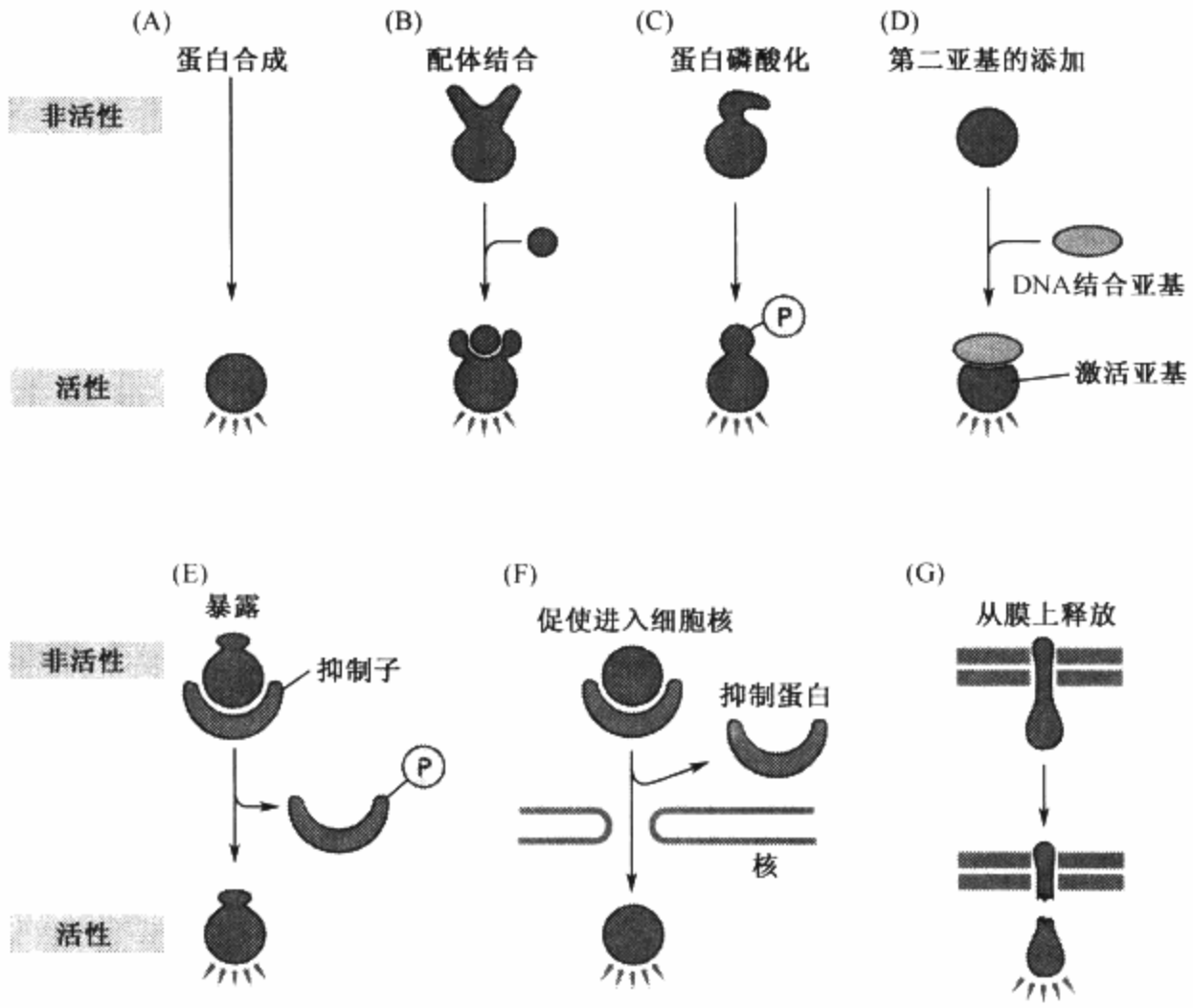


图 7-58　真核细胞中调控基因调节蛋白活性的一些方式

（A）该蛋白质只在需要的时候合成，并由蛋白质水解作用迅速降解，这样它就不会累积起来。（B）配体结合激活。（C）磷酸化激活。（D）DNA 结合蛋白和另一个具有转录激活结构域的蛋白质形成复合体。（E）抑制蛋白的磷酸化暴露激活结构域。（F）抑制蛋白的消除刺激进入细胞核，否则该抑制蛋白阻止调节蛋白进入细胞核。（G）可调控的蛋白质水解作用将基因调节蛋白从膜双层释放。这些机制每一个都典型地受细胞外信号调控，这些信号跨过细胞质膜与细胞内基因调控蛋白交流信息。这种信号传递的方式在第 15 章讨论。机制（A）～（F）容易逆转，因而提供了选择性地失活基因调节蛋白的方式

步，剩余的基因调节蛋白在 DNA 上组装起来，指导单个基因的表达。

LCR 似乎通过调控染色质压缩来发挥作用。它的重要性可以从患有地中海贫血病（贫血的一种严重的遗传形式）的患者中看出来。在这些患者中，发现 β-珠蛋白基因座发生了删除，将所有或部分 LCR 移走，虽然 β-珠蛋白基因以及它邻近的调节区域是完好的，但是整个基因即便是在类红细胞中也保持转录沉默。此外，类红细胞中的 β-珠蛋白基因不能进行类红细胞发育过程中进行的正常的染色质解压缩步骤。

人类基因组中存在许多 LCR（也就是指调节远处基因或基因簇可接触性及表达的 DNA 调节序列），它们调节了大量细胞类型特异的基因。它们发挥作用的方式还没有详细地了解，但是已经提出了许多模型。其中最简单的是基于我们本章已经讨论过的原理：结合 LCR 的调节蛋白通过 DNA 环化与结合在受它们调节的基因的调控区上的蛋白质相互作用。通过这种方式，结合在 LCR 上的蛋白质能够吸引染色质重建复合体和组蛋白修饰酶，它们能够在转录开始前改变基因座的染色质结构。有关 LCR 的其他模型提出一种机制，通过该机制最初结合在 LCR 上的蛋白质吸引其他协同组装的蛋白质，因而沿 DNA 向它们调控的基因延伸，当它们前进的时候，对染色质进行了修饰。

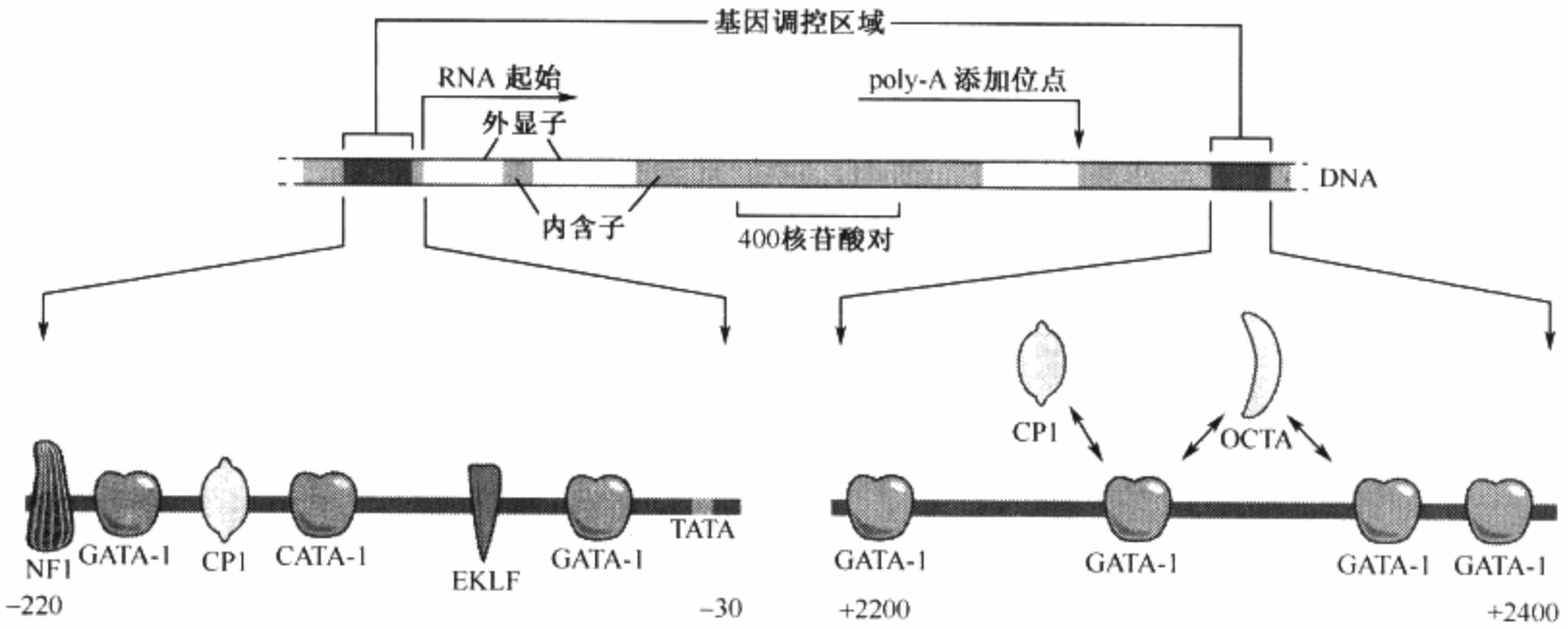


图 70-59　人类 β-珠蛋白基因的调控模型

该图显示了一些据认为在红细胞发育期间调控基因表达的基因调节蛋白（图 7-60）。所示的有些基因调节蛋白，如 CP1，在许多种细胞中都已发现，而另一些，如 GATA -1，只存在于少数几种细胞类型中（包括红细胞），因而认为有助于 β 珠蛋白基因表达的细胞类型特异性。如双箭头所示，GATA -1 的几个结合位点与其他基因调节蛋白的结合位点重叠；据认为这些位点由 GAT -1 占据时就会排斥其他蛋白质的结合。一旦结合到 DNA 上，基因调节蛋白就募集染色质重建复合体、组蛋白修饰酶、通用转录因子和 RNA 聚合酶到启动子上（摘自 B. Emerson, in Gene expression: Genral and Cell-Type Specific (M. Karin, ed.), pp. 116～161. Boston: Birkhauser, 1993）

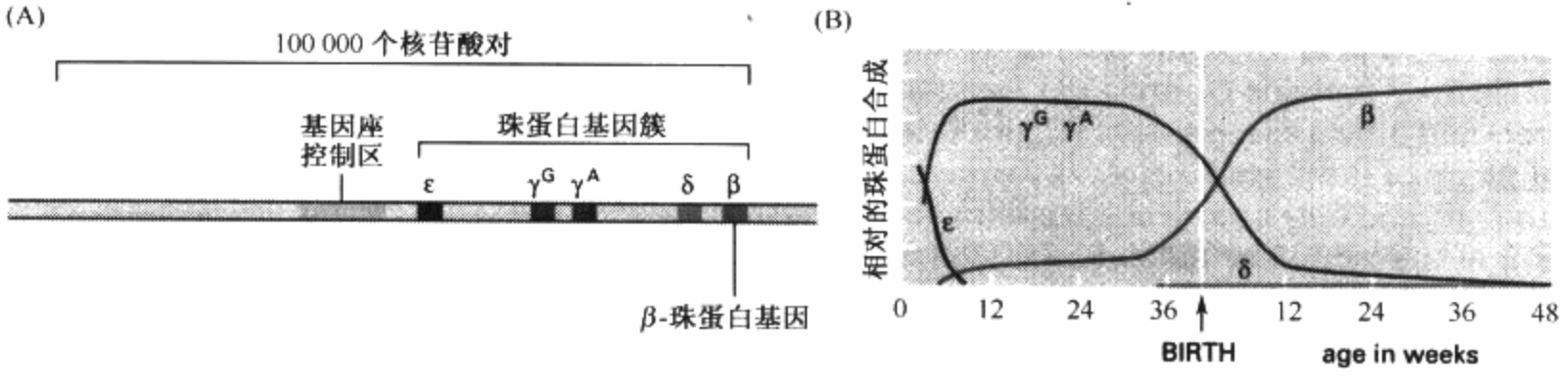


图 7-60　人的 β 类似珠蛋白基因簇

（A）图中所示巨大染色体区域跨越 100 000 个核苷酸对，含有 5 个珠蛋白基因和 1 个基因座控制区（LCR）。（B）β 类似珠蛋白基因在人发育的不同阶段表达的变化。这些基因编码的每一条珠蛋白链与一条 α-珠蛋白链组合起来形成红细胞中的血红蛋白（图 7-115）（摘自 F. Grosveld, G. B. Van Assendelft, D. R. Greaves, and G. Kollias, *Cell* 51: 975～985, 1987. © Elsevier）

绝缘子是防止真核基因调节蛋白影响远处基因的 DNA 序列

所有的基因都有调控区，它指令在什么时间，什么条件下，在什么组织中表达该基因。我们也已经看到真核基因调节蛋白可以跨越很长一段 DNA 发挥作用。那么不同基因的调控区是如何保持彼此互不干扰的呢？换句话说，是什么防止了结合在一个基因调控区的基因调节蛋白不适当地影响邻近基因的转录呢？

已经提出了许多机制来解释这种区间化，但了解得最清楚的机制依赖于绝缘子元件，也称作边界元件。绝缘子元件（简称绝缘子）是结合特殊蛋白的 DNA 序列，具有两种特性（图 7-61）。首先，它们缓冲了异染色质对基因的抑制作用。如果一个基因（如来自果蝇或者

小鼠）及其正常的调控区域插入基因组不同位置，那么根据它在基因组中插入的位点不同，它的表达水平也不同，如果插入到异染色质中，表达水平就特别低。第 4 章我们见过这种位置效应的一个例子，其中讲到插入到异染色质中的基因都转录沉默了（图 4-45）。然而，如果将该基因侧翼的绝缘子及其调控区包括在内，这个基因通常就会正常表达，无论它在基因组中的新位置如何。绝缘子的第二种特性在某种意义上与这一点相反：它们能阻断增强子的作用（图 7-61）。要产生这种效应，必须将绝缘子置于增强子和目标基因启动子之间。

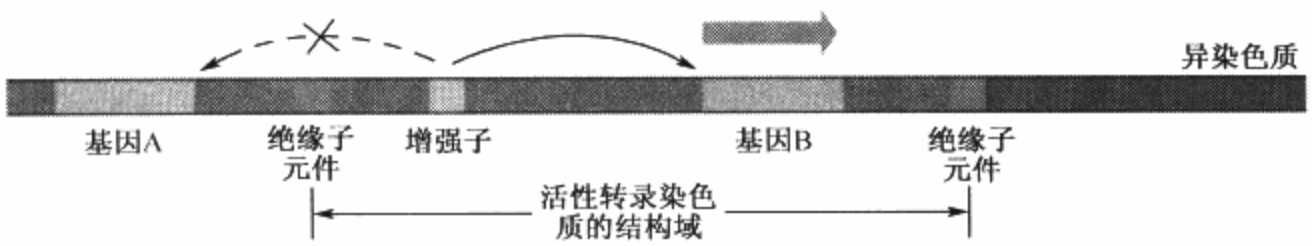


图 7-61　概括绝缘子功能的简图

绝缘子既防止异染色质（图右边）的扩散，又直接阻断增强子（左边）的作用。因此基因 B 受到适当的调节并且防止了基因 B 的增强子影响基因 A 的转录

因此绝缘子能界定基因表达的区域，既缓冲了基因外部的影响，又阻止了基因（或基因簇）的调控区对该区域外界产生影响。例如，珠蛋白 LCR（上面讨论的）与相邻的绝缘子连接起来，该绝缘子使 LCR 能够只影响这一簇珠蛋白基因。也许另外一个绝缘子位于珠蛋白基因簇的远侧，帮助定义了该区域的另一端。

因此认为绝缘子在基因组中的分布将基因组分割成独立的基因调控区域和染色质结构区域。与这种观点一致的是，绝缘子跨基因组的分布大致与染色质结构中的变化相关。例如，果蝇中的一种绝缘子结合蛋白优先定位于多线染色体的间带（以及胀泡的边缘）（图 7-62）。

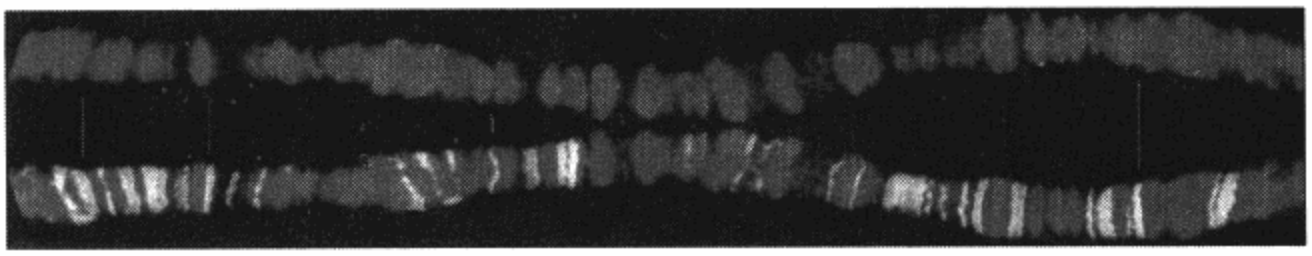


图 7-62　果蝇绝缘子结合蛋白在多线染色体上的定位

多线染色体用碘化丙锭（红色）染色以显示其带型，图中带呈亮红色，而间带呈黑色缺口（上部）。利用特定绝缘蛋白（称为 BEAF）的抗体将其在此多线染色体上结合的位置染成亮绿色（底部），可以看到 BEAF 优先位于间带区域，这表明它在组织染色质区域中发挥作用。为了方便起见，这两张显微照片以镜像形式排列（经 Uli Laemmli 许可，摘自 K. Zhao et al., *Cell* 81：879～889，1995. © Elsever）

目前还不清楚绝缘子的作用机制，可能不同的绝缘子以不同的方式起作用。至少有些绝缘子配对可能界定了成环染色体区域的基础（图 4-44）。有人提出所有的真核生物染色体都由绝缘子分割成独立的成环区域，每个都单独受到调节。

细菌利用可交换的 RNA 聚合酶亚基帮助调节基因转录

我们已经看到了基因调节蛋白的重要性，它们结合 DNA 中的调节序列，并给转录装置发出信号表示是否开始合成 RNA 链。虽然这是真核生物和原核生物中调控转录起始的主要

方式，但是有些细菌和它们的病毒采取另外的基于可互换 RNA 聚合酶亚基的策略。在第 6 章描述过，细菌 RNA 聚合酶识别启动子需要 σ 亚基。许多细菌合成多种不同的 σ 亚基，每个都能与 RNA 聚合酶核心相互作用，将它导向一组不同的启动子（表 7-2）。这种策略使得只通过替换 σ 亚基就能关闭一大组基因，而又打开另一组基因；这种策略之所以有效，是因为它绕过必须与基因一个一个作用的必要性。它常常破坏性地被病毒用于占有宿主聚合酶并迅速而且连续地活化许多套病毒基因（图 7-63）。

表 7-2　*E. coli* 的 Sigma 因子

Sigma 因子	识别的启动子
σ^{70}	大多数基因
σ^{32}	热激诱导的基因
σ^{28}	平台期和胁迫反应的基因
σ^{28}	参与迁移和趋化性的基因
σ^{54}	氮代谢基因

Sigma 因子的命名是根据它们的近似分子量(kDa)。

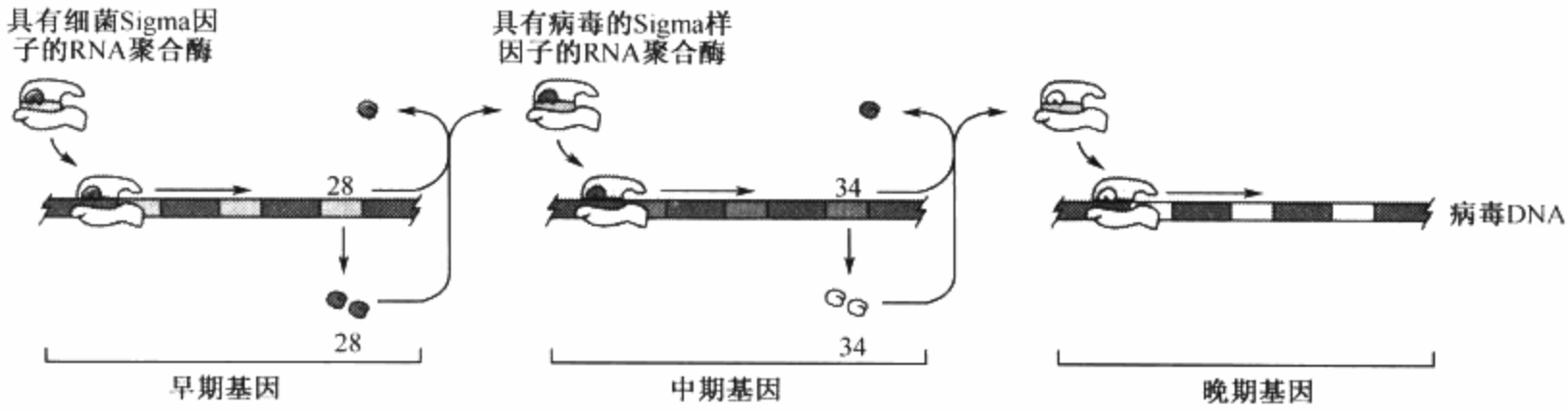


图 7-63　一种细菌病毒利用可互换的 RNA 聚合酶亚基作为一种调控基因表达的策略

细菌病毒 SPOI 感染枯草杆菌，病毒 DNA 进入细菌细胞以后不久，它就利用细菌聚合酶转录其早期基因。其中一个早期基因称为 28，编码一个 σ 类似因子，它结合 RNA 聚合酶并取代细菌的 σ 因子。这种新的聚合酶形式专门起始 SPOI “中期”基因的转录。其中一个中期基因编码另一个 σ 类似因子——34，由它取代 28 产物并指导 RNA 聚合酶转录“晚期”基因。最后表达的这套基因合成用于包装病毒染色体到病毒衣壳中并裂解细胞的蛋白质。通过这种策略，多套病毒基因以需要它们的先后顺序表达；这确保了迅速而有效的病毒复制

在某种意义上，通过利用共有一些亚基的 3 种不同 mRNA 聚合酶（Ⅰ、Ⅱ、Ⅲ），真核生物也运用了类似的策略。相比之下，原核生物只利用一种核心 RNA 聚合酶分子，但是它们用不同的 σ 亚基对它进行了修饰。

基因开关是逐渐进化而来

我们已经看到真核基因的调控区常常展开成一长段 DNA，而原核生物基因的调控区通常紧密包装在转录开始位点附近。然而，多个细菌基因调节蛋白识别远离启动子许多核苷酸对的 DNA 序列，这些我们可以见图 7-40。这种情况提供了基因调节中最早的 DNA 环化例子之一，大大影响了后来对真核基因调节蛋白的研究。

看起来似乎细菌基因开关紧密包装的排列，可能是在细菌应答进化压力来维持小基因组时，从更加伸展的开关形式发展而来的。然而这种压缩的出现是付出了代价的，因为它限制

了调控装置的复杂性和适应性。相比之下，由于真核生物调控区的伸展形式，伴之以长段间隔 DNA 分隔开的不连续调节模块，就可以期待在进化过程中促进了调节模块的重新洗牌，两者都产生新的调节环路并修改了原有的环路。基因调控区如何进化是一个令人着迷的挑战，在现今 DNA 序列中能找到许多线索。我们将在本章的最后再回到这个主题上来。

小结

细胞内单个基因的转录由基因调节蛋白打开和关闭。在原核生物中，这些蛋白质通常结合 RNA 聚合酶起始位点附近的特异 DNA 序列，并根据该调节蛋白的特性，及其相对于起始位点的精确位置，激活或者阻遏这个基因的转录。然而，DNA 螺旋的柔性也允许结合在远处位点的蛋白质通过将间隔的 DNA 环化来影响启动子上的 RNA 聚合酶。这种远距离发挥作用的情况在真核细胞内极为常见，其中远离启动子几千个核苷酸对结合的基因调节蛋白调控着基因的表达。真核激活蛋白和阻遏蛋白以多种机制发挥作用——一般引起染色质结构的局部修饰、通用转录子在启动子上的装配，以及 RNA 聚合酶的募集。

虽然一个典型的原核基因只由一个或两个基因调节蛋白调控，但是更高级的真核基因的调控要复杂得多，从而与更大的基因组以及形成了大量不同的细胞类型相称。例如，果蝇 *eve* 基因的调控区拥有长达 20 000 个核苷酸对的 DNA，以及超过 20 种基因调节蛋白的结合位点。其中一些蛋白质是转录激活蛋白，而另一些是转录阻遏蛋白。这些蛋白质与 DNA 上组织成一系列串联的调节模块的调节序列结合，并且它们共同发挥作用导致产生基因表达的空间和时间模式。真核基因和它们的调控区常常由绝缘子环绕，它们是由防止独立调节的基因之间交叉对话的蛋白质识别的 DNA 序列。

产生特化细胞类型的分子遗传学机制

尽管所有细胞都必须能够打开或关闭基因，以应答环境的变化，但是多细胞生物的细胞已经将这种能力演化到一种极端的程度，并以高度特化的方式，形成了分化细胞类型有组织的排列。尤其是，一旦多细胞生物的某个细胞开始分化成特定的一种细胞类型，这种命运的选择常常会经过许多后代细胞仍然保留着，这意味着必须记住基因表达中参与这种选择的变化。这种细胞记忆现象是形成有序组织和维持分化细胞稳定性的先决条件。相比之下，无论在真核生物还是原核生物中，基因表达发生最简单的变化都只是暂时的。例如，在细菌中色氨酸阻遏蛋白只有存在色氨酸时才将色氨酸基因关闭；一旦色氨酸从培养基中消除，基因表达再打开，并且后代细胞将不会记住它们的祖先曾接触过色氨酸。然而，即便是在细菌中，基因表达中的一些变化类型也能稳定地遗传下去。

这一部分我们考察如何将基因调节装置综合起来形成“逻辑”回路，细胞通过这些回路进行分化、计时、记住过去发生的事件，并在整个染色体水平上调节基因表达水平。首先我们讨论一些了解最清楚的细胞分化遗传机制，它们在细菌和酵母细胞中发挥作用。

DNA 重排介导了细菌的相转变

我们已经看到，在高等真核生物中进行细胞分化时，DNA 序列常常不会发生可察觉的改变。相比之下，在有些原核生物中，通过发生激活或失活特定基因的 DNA 重排，获得了一种稳定遗传的基因调节模式。由于在随后进行的 DNA 复制中 DNA 序列的变化也忠实地进行了拷贝，所以一种基因活性发生改变的状态将被发生重排细胞的所有后代继承。然而这种重

排有些可以逆转，这样，偶尔有些个体会再转回到原先的 DNA 构型。其结果就是，经过长期和许多世代的观察后，能发现基因活性的交替模式。

对这种分化机制研究得很清楚的一个例子发生在沙门氏菌中，称其为**相转变**。虽然还不知道是否在更高等的真核生物中存在相应的机制，但它对真核生物仍然能产生巨大的影响，因为致病菌利用它来逃避免疫系统的发现。一段长达 1000 个核苷酸对的特异 DNA 偶尔倒位引起沙门氏菌中基因表达的打开与关闭。这种变化改变了细胞表面蛋白——鞭毛蛋白的表达，细菌有两个不同的基因编码了这种蛋白质（图 7-64）。这种倒位由一种位点特异的酶催化，并改变了位于这 1000 个核苷酸对中一个启动子的方向。当启动子以一个方向存在时，细菌就合成一种鞭毛蛋白；当启动子以另一个方向存在时，细菌就合成另一种鞭毛蛋白。由于发生倒位的情况很罕见，所以长出来的细菌全部克隆只会具有其中一种类型的鞭毛蛋白。

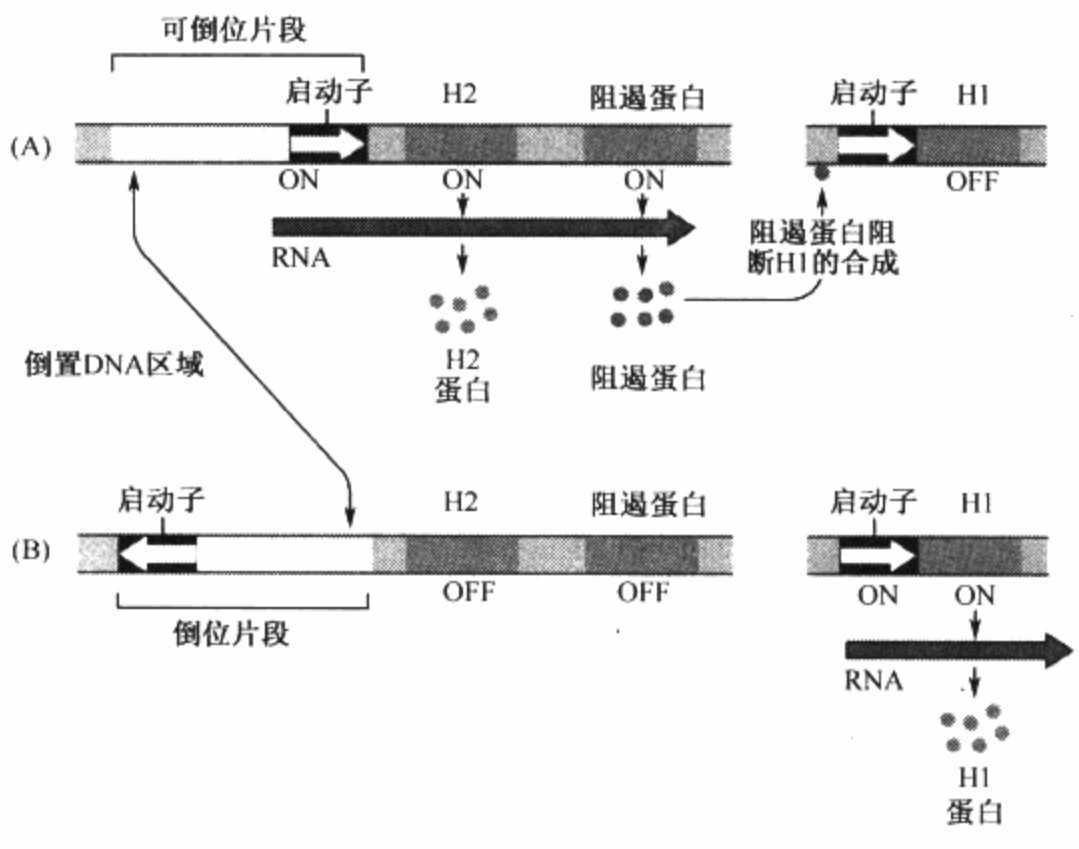


图 7-64　细菌中 DNA 倒位改变基因表达

一种简单的位点特异性重组使含有启动子的小 DNA 片段发生倒位，从而引起沙门氏菌中两个鞭毛蛋白基因交替转录。（A）以一种方向存在时，启动子激活 H2 鞭毛蛋白基因的转录，还激活了一种阻断 H1 鞭毛蛋白基因表达的阻遏蛋白的转录。（B）启动子发生倒位以后，它不再打开 H2 或者阻遏蛋白，因此 H1 从阻遏状态释放出来并进行表达。重组机制很少得到激活（大约每 10^5 次细胞分裂出现一次）。因此在每个细胞克隆中这种或另一种鞭毛蛋白的合成倾向于忠实地遗传

相转变几乎必定会进化，因为保护了细菌群落免受其脊椎动物宿主免疫反应的伤害。如果宿主产生针对一种鞭毛蛋白的抗体，那么少数鞭毛蛋白因基因倒位而发生改变的细菌仍将能够存活和增殖。

分离的野生型细菌经常显示出一种或多种表型的相转变。这些“不稳定性”常常随着时间从标准实验室菌株中丢失，其中的机制还仅仅在少数例子中研究过。不是所有情况都涉及 DNA 倒位。例如，有一种引起一类常见的靠性传播的人类疾病的细菌（淋球菌），通过其表面特征发生一种可遗传的改变的方式逃避免疫攻击。这种改变的发生，是因为基因转变（讨论见第 5 章）而不是基因倒位。这种机制将 DNA 序列从沉默“基因盒”文库中转移到基因组中基因进行表达的位点；其优势就是能产生主要细菌表面蛋白的许多变体。

一组基因调节蛋白决定了芽殖酵母的细胞类型

由于酵母很容易培养和进行遗传操作，所以它已作为研究真核细胞中基因调控机制的模式生物。常见的面包酵母、酿酒酵母，由于能分化成 3 种不同的细胞类型，因而特别吸引研究者的兴趣。酿酒酵母是一种单细胞真核生物，以单倍体或二倍体状态存在。二倍体细胞由两个单倍体细胞通过称作**交配**的过程融合而成。两个单倍体细胞为了能够交配，它们的交配型（性别）必须不同。在酿酒酵母中存在两种交配型 a 和 α，它们专门用于与对方交配。它们每个分别产生一种特殊的可扩散的信号分子（交配因子）和一种特殊的细胞表面受体蛋白。它们联合起来使细胞能够与相反的细胞类型互相识别，然后同它进行融合。结果形成的二倍体细胞称为 a/α，与父母中的任何一个都不相同：当它们耗尽所有养分时，它们不能交配但能形成孢子（孢子形成），从而通过减数分裂产生单倍体细胞（讨论见第 20 章）。

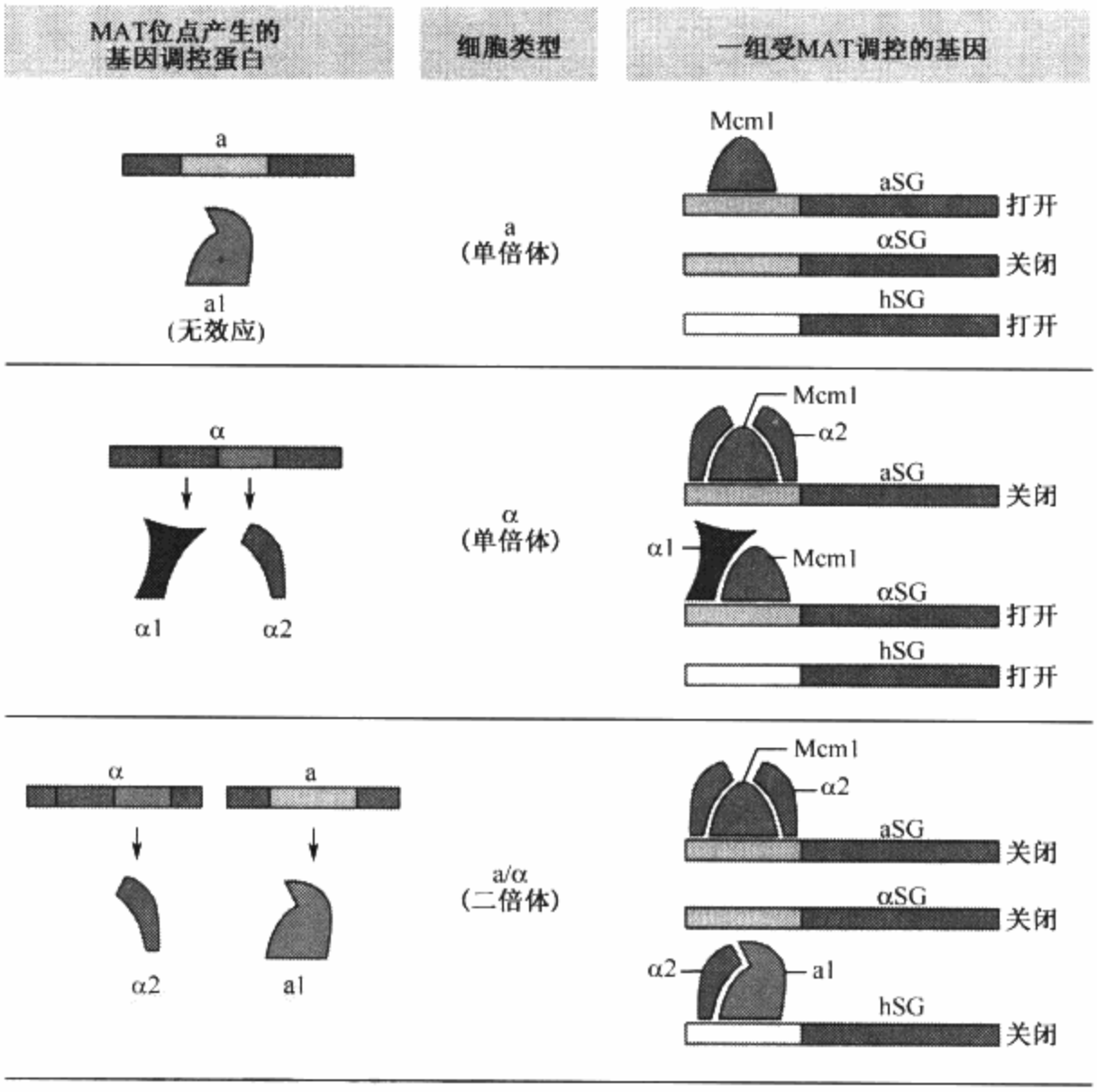


图 7-65　酵母细胞类型的调节

酵母细胞类型由 MAT 基因座产生的 3 种基因调节蛋白（α1、α2、a1）决定。a 型单倍体细胞、α 型单倍体细胞和二倍体细胞（a/α 型）中分别转录一套不同的基因。单倍体细胞表达一套单倍体特异的基因（hSG），既可以是一套 α 特异基因（αSG），也可以是一套 a 特异基因（aSG）。二倍体细胞不表达这些基因。α1、α2 和 a1 蛋白通过以不同的组合方式，结合每种细胞类型中许多目标基因上游特异的调节序列，从而调控这些基因的表达。注意 α1 是一个基因激活蛋白，而 α2 是一个基因阻遏蛋白。二者都与称作 Mcm1 的基因调节蛋白组合起来发挥功能，这种蛋白在全部 3 种细胞类型中都存在。在二倍体细胞类型中，α2 和 a1 形成一个异源二聚体（图 7-23 详细显示），由它关闭一套基因（包括编码 α1 激活蛋白的基因）的表达，这套基因与 α2 和 Mcm1 蛋白关闭的基因不同。这个相对简单的基因调节蛋白系统是基因表达组合调控的一个实例（图 7-50）。a1 和 α2 蛋白质各自利用一个同源异型域基序识别它们 DNA 结合位点（图 7-16）

建立和维持这 3 种细胞类型的机制说明了我们已经讨论过的改变基因表达模式的几个策略。单倍体细胞的交配型由单个位点，**交配型（MAT）位点**决定，它在 a 型细胞中编码一个基因调节蛋白——a1；在 α 细胞中编码两个基因调节蛋白——Matα1 和 Matα2。a1 蛋白在产生它的 a 型单倍体细胞中无任何作用，但是后来在交配产生的二倍体细胞中变得重要起来；同样，α 型单倍体细胞不会产生其交配型特异的蛋白质。相比之下，α2 蛋白在 α 细胞中作为一种转录阻遏蛋白将 a 特异的基因关闭，而 α1 蛋白则作为转录活化蛋白将 α 特异的基因打开。一旦这两种交配型发生融合，a1 与 α2 调节蛋白的组合就生成一种全新的基因表达模式，与父母细胞的任何一个都不同。图 7-65 说明了交配型特异的基因在 3 种细胞类型中以不同模式表达的机制，并仍然是最先鉴定出的组合基因调控例子之一。

虽然在大多数酿酒酵母的实验菌株中，a 和 α 细胞类型经过多次细胞分裂仍保持稳定，但是一些分离得到的野生型菌株能够通过基因重排机制反复在 a 和 α 细胞类型之间转换，这种机制的作用让人想起在淋球菌中发生的 DNA 重排，尽管确切的机制似乎为酵母所特有。在酵母染色体中的 MAT 位点两侧，各存在一个沉默的位点，编码了交配型基因调节蛋白：一侧的沉默位点编码 α1 和 α2；另一侧编码 a1。几乎每次进行其他细胞分裂时，MAT 位点中的活性基因被切除，并由决定相反交配型的沉默位点新合成的拷贝替换。由于这种改变涉及将一个基因从活性狭槽中移走，并由另一个取代，所以这种机制称为弹夹机制。这种改变可以逆转，因为尽管去除了 MAT 位点原先的基因，基因组中仍然保留着一个沉默拷贝。从这个沉默的基因产生的新拷贝作为可丢弃的盒带发挥作用，它将会交替插入到 MAT 位点，就像“播放磁头”一样起作用（图 7-66）。

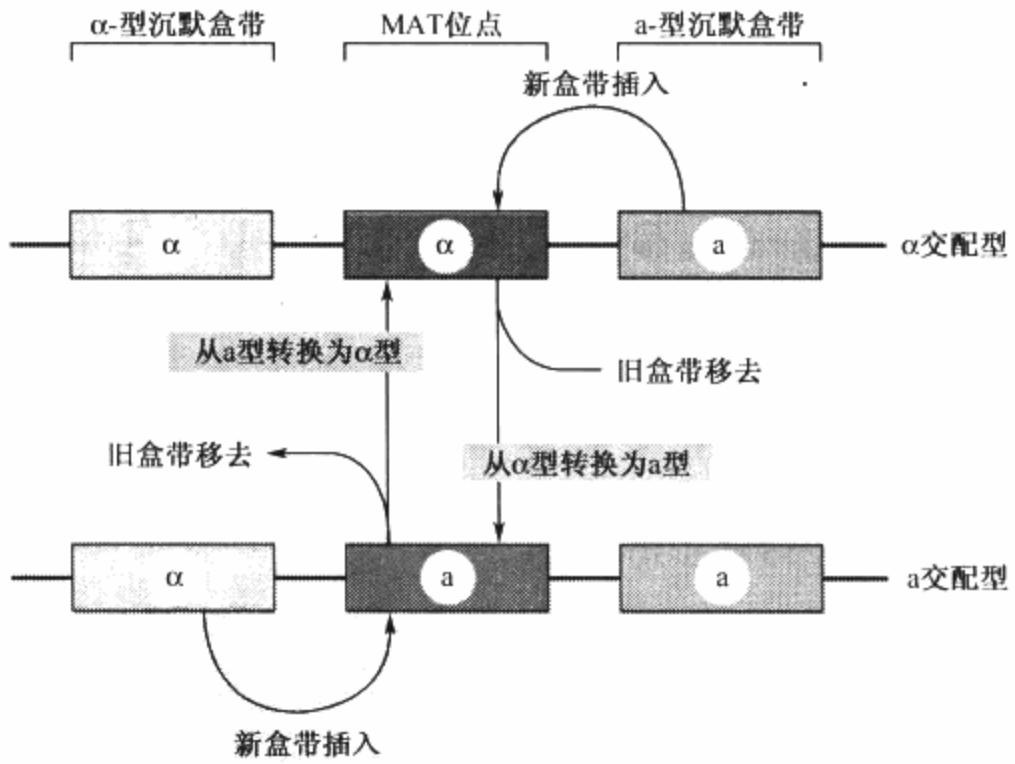


图 7-66　酵母交配型转变的弹夹模型

弹夹式转变通过一种基因转变过程发生，其中涉及一种特殊的酶（HO 内切核酸酶），它在 MAT 基因座中特殊的 DNA 序列上切开。然后将切口附近的 DNA 切除，并用相反交配型的沉默弹夹拷贝取代。这种特殊形式的基因转变机制与第 5 章讨论的同源末端连接的一般机制相似

负责沉默位于酵母染色体末端基因的相同机制，以一种无转录活性的形式维持着沉默盒（图 4-47）；也就是说，位于沉默位点的 DNA 包装成一种高度组织的染色质形式，抑制基因的表达。

两种彼此阻遏对方合成的蛋白决定了λ噬菌体的遗传状态

单个体细胞核内存在的遗传信息能够指令完整的脊椎动物或植物个体（图 7-2）的事实，消除了 DNA 序列的不可逆变化成为高等真核细胞主要分化机制的可能性（虽然这种改变是淋巴细胞分化中的关键；讨论见第 24 章）。从理论上讲，可逆的 DNA 序列变化，与刚刚描述的在淋球菌和酵母中发生的那些变化类似，仍然可以负责产生在高等生物基因表达中观察到的遗传变化，但是目前没有证据表明这种机制已得到广泛利用。

然而，本章我们曾谈及的其他机制也能产生可由后代细胞继承的基因调节模式。可能最简单的例子存在于λ细菌病毒（λ噬菌体）中。在λ细菌病毒中存在一个开关，能引起该病毒在两种稳定的自我维持状态之间转变。这种开关可以看作是在高等真核生物发育中发挥作用的类似但更复杂的开关的原型。

我们在前面提到过λ噬菌体在有利的条件下能整合到大肠杆菌细胞 DNA 中，以便每次细菌分裂时都自动得到复制。该病毒也可以在细胞质中增殖，将宿主杀死（图 5-81）。这两种状态之间的转换由噬菌体基因组编码的蛋白质介导。其基因组含有大约 50 个基因，在两种状态下它们以非常不同的模式转录。例如，一个注定要整合的病毒必须合成λ整合酶蛋白，需要利用它将 λDNA 插入细菌染色体，但必须阻遏负责病毒增殖的病毒蛋白的合成，一旦这两种模式之一确立了，它就会稳定地维持下去。

我们这里不能详细地讨论这个复杂的基因调控系统，而只是概括一些它的一般特征。这种转换的核心是该病毒合成的两种基因调节蛋白：**λ阻遏蛋白**（CI 蛋白）（我们已经涉及过）和 **Cro 蛋白**。这两种蛋白质彼此抑制对方的合成，这种安排正好产生两种稳定状态（图 7-67）。在状态 1（原噬菌体状态）中，λ阻遏蛋白占据启动子，阻断 Cro 的合成，也激活它自身的合成；在状态 2（裂解状态）中，Cro 蛋白占据操纵子中另一个位点，阻断阻遏蛋白的合成，但允许它自身进行合成。在原噬菌体状态中，大部分稳定整合的噬菌体 DNA 不进行转录。在裂解状态，该 DNA 大量地转录、复制，并包装成新的噬菌体，由宿主细胞裂解释放。

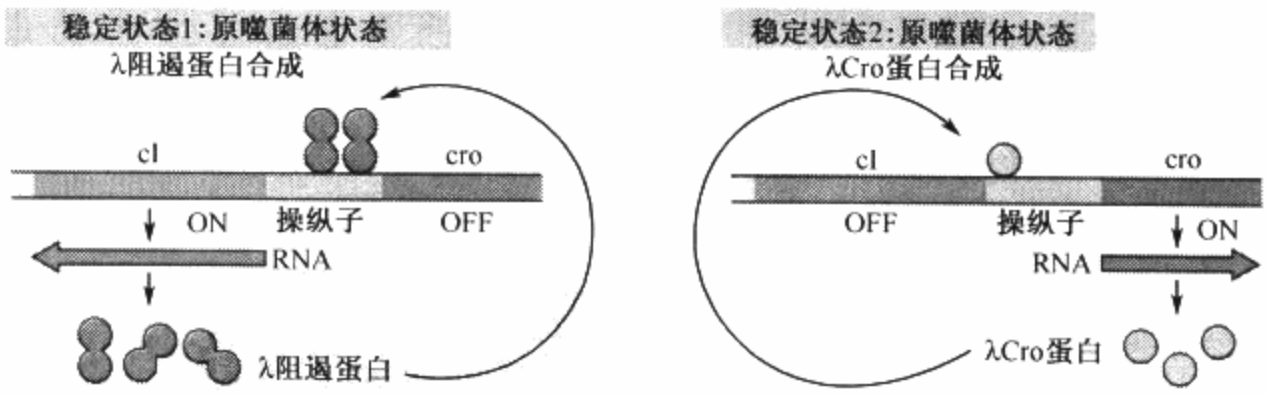


图 7-67　大肠杆菌宿主细胞内，决定λ噬菌体生长模式的调节系统的简化形式

在稳定状态 1（原噬菌体状态）中，噬菌体合成一种阻遏蛋白，它激活自身的合成，并关闭多种其他噬菌体蛋白质的合成，其中包括 Cro 蛋白。在状态 2（裂解状态）中，噬菌体合成 Cro 蛋白，它关闭掉阻遏蛋白的合成，这样就能合成许多噬菌体蛋白质，而且病毒 DNA 能在大肠杆菌细胞内自由复制，最终产生了许多新的噬菌体颗粒并杀死细胞。该例子说明了两个基因调控蛋白如何能够组合成一条回路来产生两种可遗传状态。λ阻遏蛋白和 Cro 蛋白都通过螺旋-转角-螺旋基序来识别操纵基因（图 7-14）

当宿主菌生长良好时，感染病毒倾向于采用状态 1，允许病毒 DNA 随宿主染色体一起增殖；当宿主受到损伤时，为了在细胞质中增殖并迅速逃离宿主，整合的病毒就从状态 1 转变

到状态 2。这种转变在宿主应答使阻遏蛋白失活的 DNA 损伤时引发。然而，如果没有这种干扰，λ 阻遏蛋白不仅关闭 Cro 蛋白的合成，而且打开自身的合成，这种正反馈环帮助维持着原噬菌体状态。

基因调节环路能用于组建记忆装置以及振荡器

正反馈环为细胞记忆提供了一种简单的通用策略——确立和维持基因转录的可遗传模式。图 7-68 显示了最核心的基本原理，真核细胞广泛地利用了这种简单策略的变化形式。例如，参与建立果蝇身体建成（讨论见第 21 章）的多种基因调节蛋白，激活它们自身的转录，因而形成一个正反馈环，促进它们继续合成；同时，这些蛋白质有许多阻遏了编码其他重要基因调节蛋白基因的转录。通过这种方式，只需要几个基因调节蛋白相互影响各自的合成和活性，就能获得遗传行为的复杂模式。

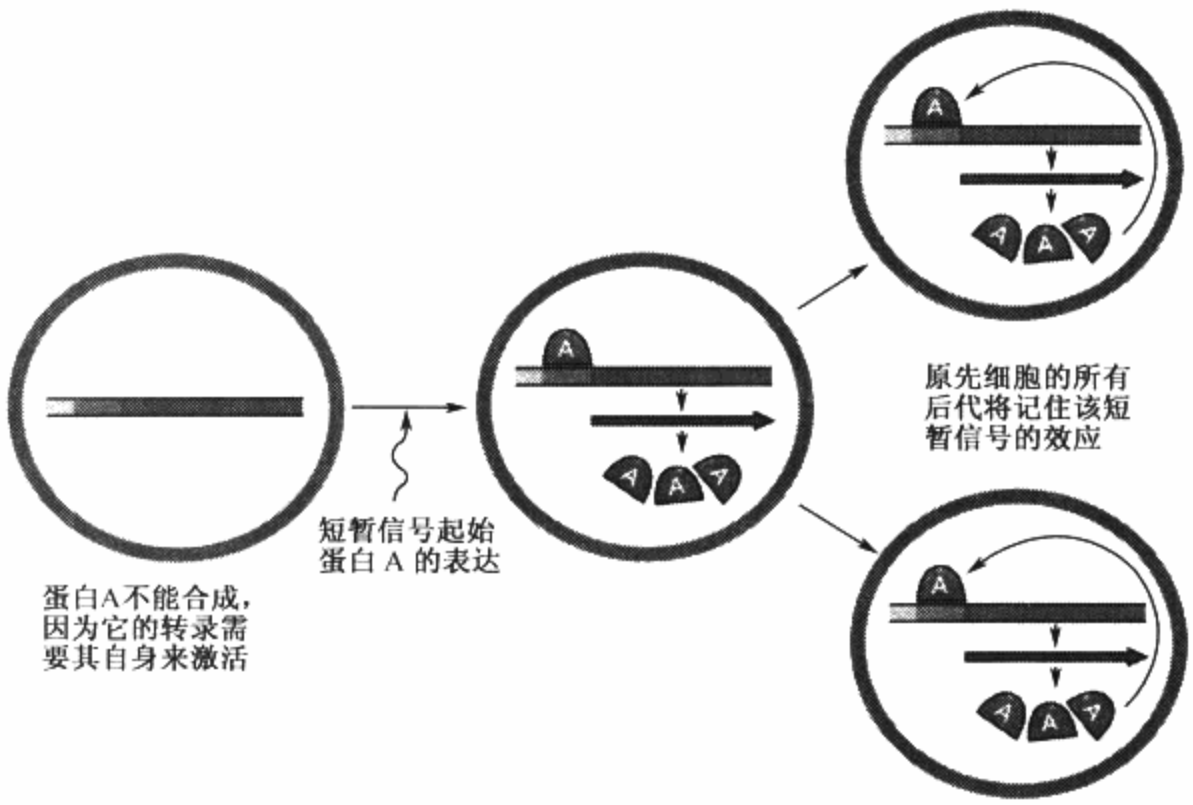


图 7-68　显示正反馈环如何产生细胞记忆的简图

蛋白 A 是一种基因调节蛋白，能激活自身的转录。原先的细胞所有后代因而将记住祖先细胞经历过起始该蛋白质合成的短暂信号

简单的调节环路可以组合起来形成多种多样的调控装置，就像电脑中简单的电子转换元件组合起来行使各种复杂的逻辑运算一样。我们已经看到，λ 噬菌体提供了一个能在两种稳定状态之间转换的环路例子。更加复杂的调节网络类型不仅能在自然界找到，而且也能在实验室中进行设计和构建。例如，图 7-69 显示了一个工程细菌细胞如何能够在 3 种状态之间按规定的顺序转换，因而像一个振荡器或者“时钟”一样发挥作用。

昼夜节律钟以基因调节中的反馈环为基础

每天白昼与黑夜循环往复的同时，地球上的生命在不断进化。许多现今生物（从古细菌到植物再到人类）已经拥有了一种内在的节律，指令一天中在不同时间发生的不同行为。这些行为多种多样，可以是真菌的代谢酶活性的循环变化，也可以是人类精确的睡眠-苏醒循环。调控这种昼间节律的内在振荡器称为昼夜节律钟。

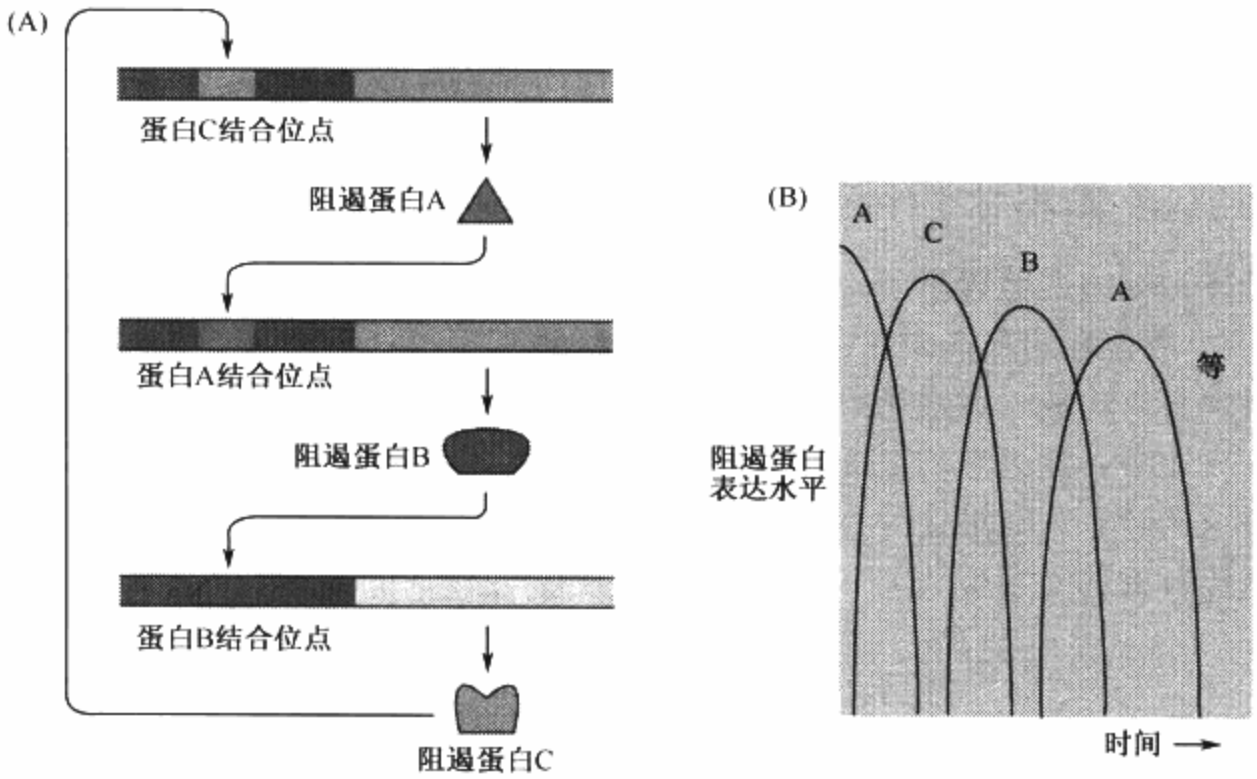


图 7-69　在实验室内设计的一种简单的基因钟

（A）利用重组 DNA 技术将 3 种不同细菌阻遏蛋白的基因分别置于一个不同的阻遏蛋白调控之下。这些阻遏蛋白（图中用 A、B、C 表示）分别是 Lac 阻遏蛋白（图 7-38）、λ 阻遏蛋白（图 7-67）和 tet 阻遏蛋白，它在应答四环素信号时调节基因表达。（B）细胞群将在（A）中显示的 3 种不同状态的循环。例如，如果细胞开始于仅有阻遏蛋白 A 累积到较高水平的状态，那么阻遏蛋白 B 就受到完全的阻遏。当阻遏蛋白 C 逐渐合成时，它开始关闭阻遏蛋白 A 的合成，于是阻遏蛋白 B 开始累积并最终关闭阻遏蛋白 C 的合成。随着此循环继续向前进行，同步的细胞群就以规定的秩序在 3 种状态之间振荡

通过携带其自身的昼夜节律钟，生物体能预测所处环境中有规律的每日变化，并提前采取行动。当然，固有的节律钟不可能绝对地准确，所以它必须能由外部信号如白天光照进行重新设定。因此即便在环境信号（光照和黑暗的变化）已不存在的情况下，昼夜节律钟也会继续运转。但是这个自由运转节律的周期常常比 24h 稍长或者稍短。表示白昼时间的外部信号引起时钟运转进行微调，以便使生物体保持与其环境同步。如果发生更激烈的改变，新的光照与黑暗循环就会逐渐重新设定（产生）昼夜节律钟，每个经历过飞行时差反应的人都能证明这一点。

有人可能推想在复杂的多细胞生物如人类中，昼夜节律钟本身会是一个复杂的多细胞装置，不同的一组细胞负责振荡机制的不同部分。然而引人注意的是，实际上几乎所有生物，包括人类的计时员都是单个细胞。因此，我们的睡眠与苏醒循环、体温的循环，以及激素释放的昼夜循环由下丘脑（脑的一部分）一组特殊细胞的每个成员（SCN 细胞）中运转的节律钟控制。即便将这些细胞从大脑中去除，分散到培养平板上，它们也将会独自继续振荡，显示出周期约为 24h 的基因表达循环模式。在完好的体内，SCN 细胞接受来自视网膜的神经信号，使它们产生光照和黑暗的日循环，并且它们将有关每天时间的信息传递到其他的组织如松果腺中，该腺体随着节律钟的时间释放褪黑素激素，将时间信号呈递到身体的其余部位。

虽然 SCN 细胞在哺乳动物中作为计时员发挥了主要的作用，但是已有证据表明它们不是哺乳动物体内唯一拥有内在昼夜节律或根据光照重新设定该钟能力的细胞。同样，在果蝇中，许多不同类型的细胞，包括胸部、腹部、触角、腿、翅膀和精巢的细胞，当与果蝇身体其余

部分分开之后，全部都持续进行昼夜循环。这些分离组织中的节律钟，与 SCN 细胞中的一样，能够被外界强加的光照和黑暗循环重新设定。

因此，昼夜节律钟的运行机制成为细胞生物学中主要难题。虽然我们还不清楚全部的细节，但在大量不同生物中进行的研究已经揭示了许多基本原理和分子组分。对动物来说，我们所知的很多都来自于对果蝇突变体的搜寻，这些突变体中的昼夜节律钟或走快或走慢或根本不走。这类工作导致发现了许多相同的组分参与了哺乳动物昼夜节律钟。图 7-70 概括了果蝇节律钟的机制。振荡器的核心是转录反馈环，它带有一种时间延迟性：某些重要基因产物的积累关闭它们的转录，但延迟转录，所以（粗略地讲）细胞在存在产物却转录关闭与不存在产物却转录打开这两个状态之间转换。

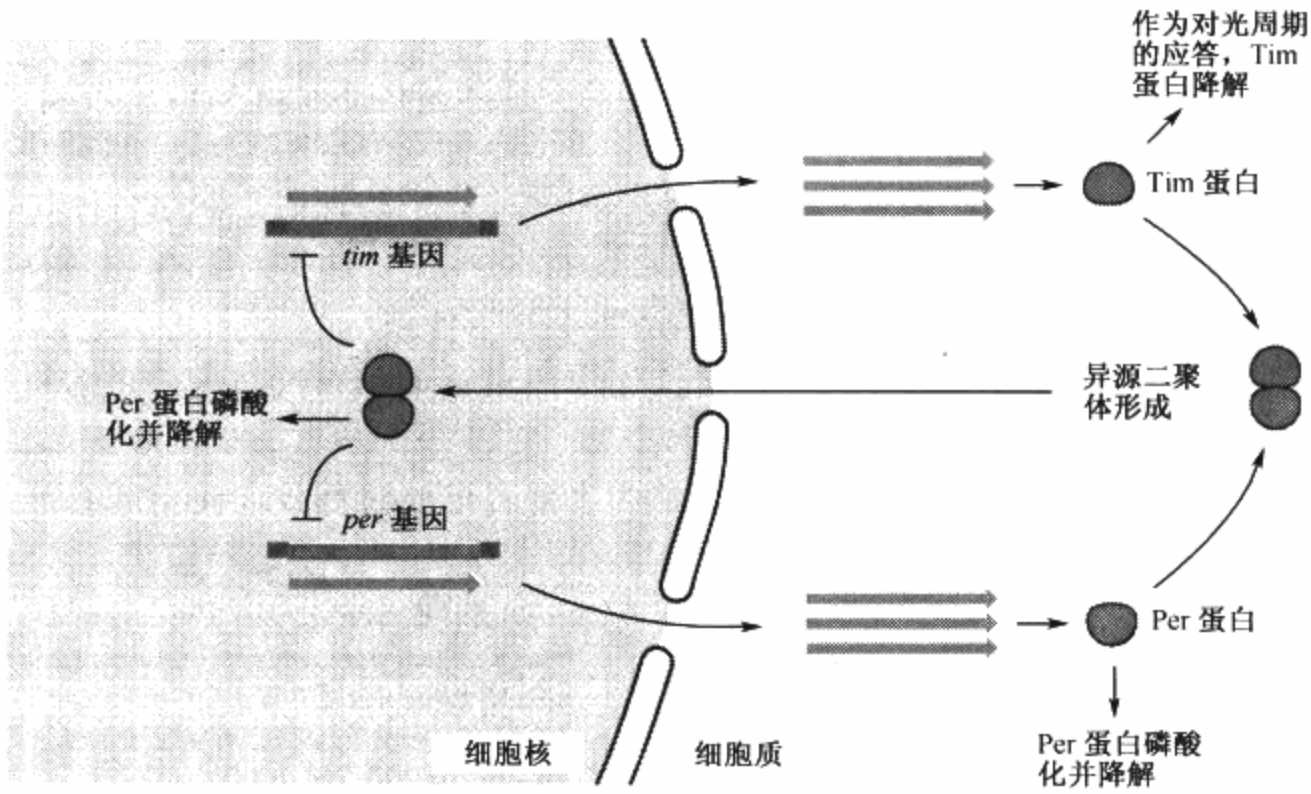


图 7-70　果蝇细胞中昼夜节律钟机制的要点

该节律钟的主要特征就是两种基因调节蛋白，Tim（timless 的简写，基于一种基因突变的表型）和 Per（period 的简写），周期性地累积和衰减。这些蛋白质在细胞质中表达，当它们累积到临界水平时，就会形成异源二聚体。该异源二聚体转运到细胞核内，在这里与该节律钟一起调节大量基因的表达。Tim-Per 异源二聚体也阻遏 *tim* 和 *per* 基因，形成了一个反馈系统，引起 Tim 和 Per 水平周期性地升高和下降。除此转录反馈之外，该节律钟依赖 Per 蛋白的磷酸化和随后的降解，这在细胞核和细胞质都有发生，且受另外一种节律钟 Dbt（double-time 的简写）的调节。降解迫使 Tim 和 Per 周期性的累积发生延迟，这对于节律钟发挥功能是很关键的。例如，Per 在细胞质中的累积因游离的 Per 单体发生磷酸化和降解而延迟。红色显示迫使发生特定延迟的步骤。对新的昼夜周期做出应答会发生节律钟的内外偶联（或者重调）。虽然大多数果蝇细胞没有真正的光受体，但细胞内的黄素蛋白感受到光，并迅速引起 Tim 降解，因而重调该节律钟

虽然昼夜节律钟的基本原理比较简单，但细节很复杂。产生这种复杂性的一个原因是，这个节律钟必须缓冲通常加速或减慢了大分子结合的温度变化。它们还必须精确地运行但是要能够被重新设定。尽管尚不清楚生物钟如何以连续的速度运行，而不论温度怎样变化，但是光诱导破坏了一种重要的基因调节蛋白是重新设定果蝇节律钟的机制，如图 7-70 所示。

单个蛋白质可以协同一组基因的表达

细胞需要能单独打开或关闭基因，但它们也需要协同大量不同基因的表达。例如，当一个静止的真核细胞接收到要进行分裂的信号时，许多到这时为此没有表达的基因就一起打开，启动最终导致细胞分裂的事件（讨论见第 18 章）。细菌协同一组基因表达的方式之一就是将它们串联起来组成一个操纵子，处于单个启动子的调控之下（图 7-33）。然而在真核生物中每个基因从单独的一个启动子开始转录。

真核生物如何协同基因表达呢？这是一个特别重要的问题，因为我们已经看到过，大部分真核基因调节蛋白作为调节蛋白“委员会”的成员之一发挥作用，它们对于在正确的细胞，正确的时间，应答适当的信号，以适当的水平表达基因全部都是必需的。那么真核细胞如何能够迅速而果断地打开或关闭哪一组基因呢？答案在于即使基因表达调控是组合型的，单个基因调节蛋白的作用，对于打开或关闭任何特定基因还是起决定性作用的，仅仅是通过完成最大限度地激活或阻遏那个基因所需的组合。这种情况与拨密码锁最后一个数字时类似：如果之前已拨入其他数字，密码锁就会弹开。正如相同的数字能完成不同组合的密码锁一样，同一蛋白质也能完成多个不同基因调控组合。如果许多不同的基因含有相同基因调节蛋白的调节位点，那么该蛋白质就可以用于调节所有这些基因的表达。

人类中这样的一个例子是由糖皮质激素受体蛋白进行的基因表达调控。为了结合 DNA 中的调节位点，这个基因调节蛋白必须首先与一分子糖皮质固醇激素如皮质醇形成一个复合物（图 15-12 和图 15-13）。这种激素在挨饿或机体剧烈运动时释放，在它的其他的活性中，它还刺激肝细胞增加从氨基酸和其他小分子合成葡萄糖。为了产生这种应答，肝细胞增加许多不同基因的表达，它们编码了代谢酶或其他产物。虽然所有这些基因有不同而且复杂的调控区，但是它们最大表达量决定于激素-糖皮质受体复合物与对每个基因 DNA 中的调节位点的结合。当身体已经恢复，激素不再存在的时候，肝中的这些基因的表达各自都降低到正常水平。通过这种方式，单个基因调节蛋白能调控许多不同基因的表达（图 7-71）。

糖皮质激素受体的作用并不只限于肝细胞。在其他细胞类型中，激素引起这个基因调节蛋白的活化，也会导致许多基因表达水平发生改变；然而受影响的基因通常与在肝细胞中受影响的基因不同。我们已经看到，每种细胞类型都拥有一套适合自身的调节蛋白，并且由于组合调控的存在，这些蛋白质精细地调控着糖皮质激素受体的作用。由于该受体能够与许多组不同的细胞类型特异的基因调节蛋白组装，所以它能在不同细胞类型中产生一系列不同的效应。

一个关键基因调节蛋白的表达能够引发一整串下游基因的表达

协同性地打开或关闭许多基因的能力不仅在日常调节细胞功能时很重要，它也是真核细胞在胚胎发育期间分化成特殊的细胞类型的方式。肌肉细胞的发育提供了一个显著的例子。

哺乳动物骨骼肌细胞是一种高度与众不同的巨大细胞，由许多称为成肌细胞的肌前体细胞融合后形成，因而含有多个细胞核。成熟的肌细胞可以通过许多独特的蛋白质与其他细胞相区别，包括肌动蛋白、肌球蛋白、原肌球蛋白和肌钙蛋白（收缩装置的全部组分）、肌酸磷酸激酶（用于肌肉细胞特殊的代谢）、乙酰胆碱受体（用于使膜对神经刺激敏感）。在增殖成肌细胞时，或者会不存在这些肌肉特异的蛋白质以及它们的 mRNA，或者以很低的浓度存在。当成肌细胞开始彼此融合时，相应的基因全部协同性地打开，作为基因表达模式全面转变的一部分。

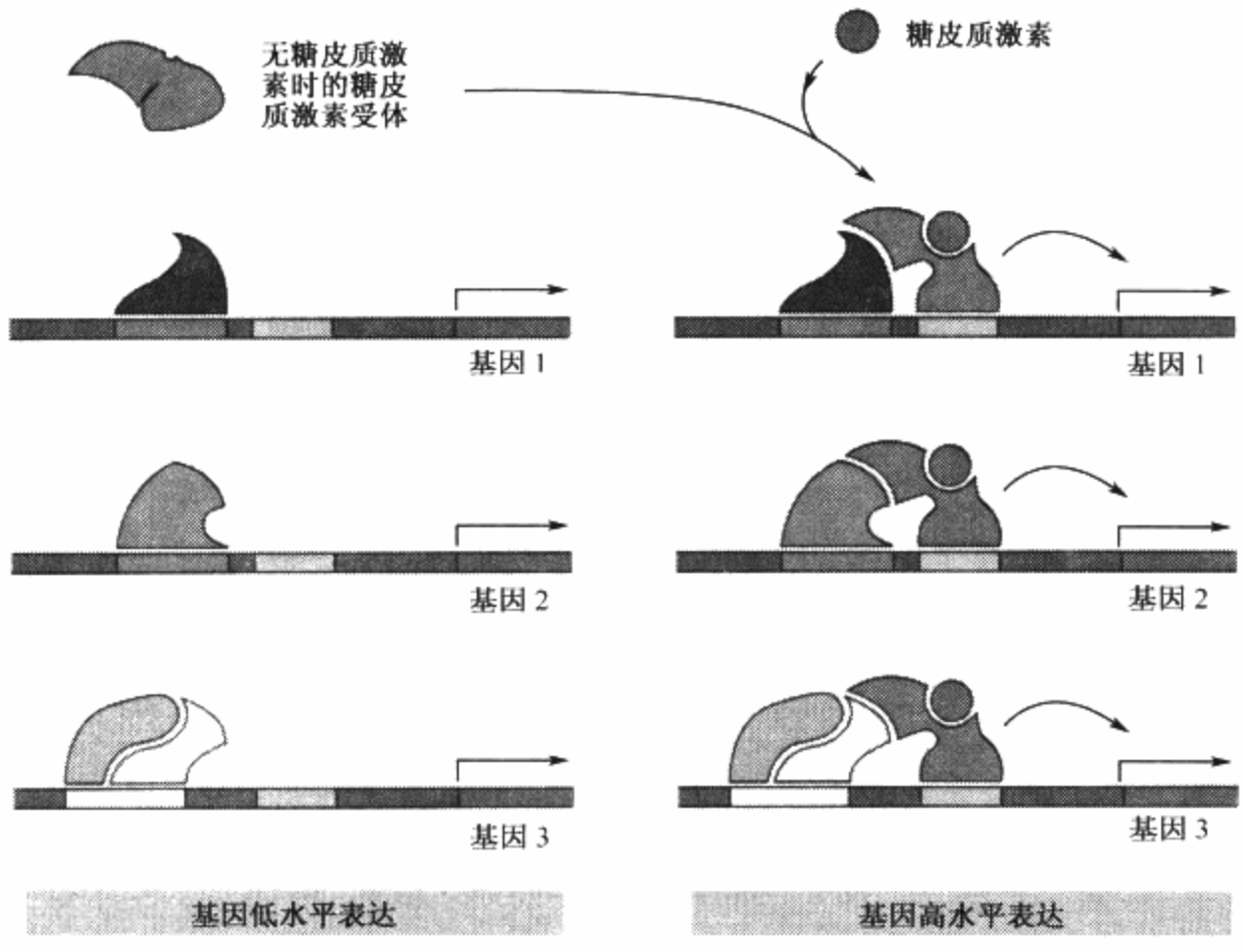


图 7-71　单个基因调节蛋白能协调几个不同基因的表达

图中概括说明了糖皮质激素受体的作用。左边的是几个基因，它们每个都有不同的基因调节蛋白结合在其调控区上。然而，单独这些结合蛋白不足以完全激活转录。右边显示的是增加另一个基因调节蛋白后（结合有糖皮质激素的复合体中的糖皮质激素受体）的效果，该蛋白质能结合每个基因的调控区。糖皮质激素受体完成了转录最大起始所需的基因调节蛋白的组合，现在这些基因就成套地打开。缺少激素时，糖皮质激素受体就保留在细胞质内，因而不能用于结合 DNA。除激活基因表达以外，激素结合形式的糖皮质激素受体阻遏某些基因的转录，这依赖于它们调控区已经存在的基因调节蛋白。糖皮质激素受体对任何特定的基因的作用因而决定于细胞类型，其所含基因调节蛋白以及基因调控区。图 7-19 显示了糖皮质激素受体 DNA 结合部分的结构

通过引入通常只在肌细胞中表达的螺旋-环-螺旋蛋白家族的任何一个成员——所谓的生肌蛋白（MyoD、Myf5、肌细胞生成蛋白和 Mrf4），可以在培养的皮肤成纤维细胞和某些其他的细胞类型中引发肌纤维分化的整个程序[图7-72(A)]。这些调节蛋白的结合位点存在于邻近的许多肌肉特异基因的调节 DNA 序列中，生肌细胞因而直接活化了许多肌肉特异的结构基因的转录。此外，生肌细胞刺激它们自身的转录以及各种参与肌肉发育的其他基因调节蛋白的转录，形成了一系列复杂的正反馈环，放大并维持着肌肉发育程序，即便起始信号已经消失以后[图7-72(B)，也可见第22章]。

也可能成纤维细胞和通过加入成肌细胞转变成肌细胞的其他细胞类型已经累积了许多基因调节蛋白，它们能与肌细胞生成蛋白质共同作用打开肌肉特异的基因。在这个观点中，是基因调节蛋白的特定组合而不是单个蛋白质决定了肌肉的分化。这个观点与某些细胞类型不能由肌细胞生成蛋白质或其相关物转变成肌肉的事实是一致的；这些细胞可能没有积累所需的其他基因调节蛋白。

通过单个基因调节蛋白的作用，一种细胞类型（成纤维细胞）转变成为另一种类型（骨骼肌细胞），这再次强调了本章讨论的最重要的原理之一：基因表达差异能产生细胞类型之间的显著差别——大小、形状、化学和功能。

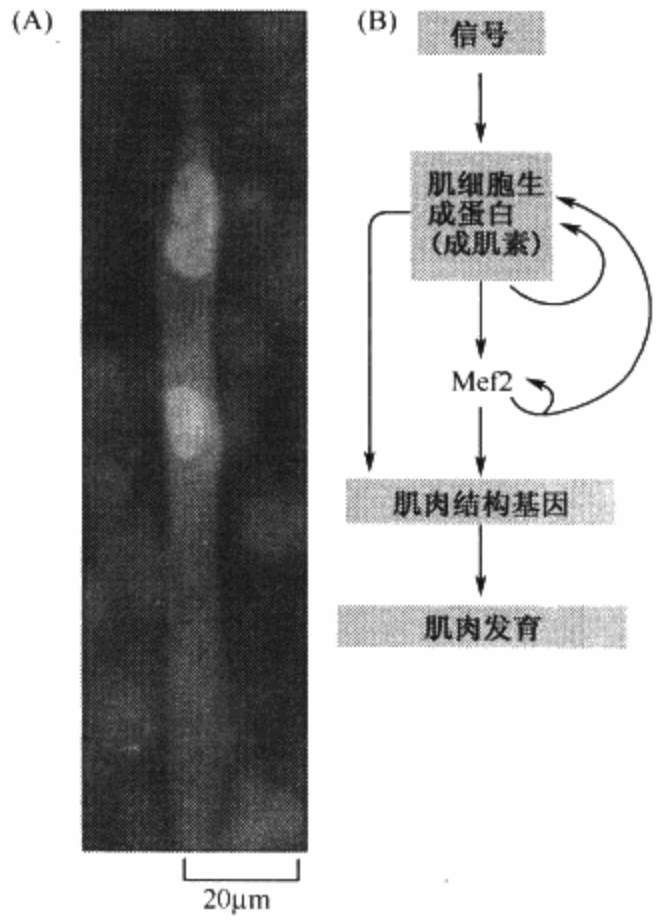


图 7-72　生肌调节蛋白在肌肉发育过程中的作用

（A）在成纤维细胞中表达 MyoD 蛋白的效果。如该免疫荧光显微照片中所示，通过实验手段诱导 *myoD* 基因的表达，将来自雏鸡胚胎皮肤的成纤维细胞转化成肌肉细胞。经诱导表达 *myoD* 基因的成纤维细胞发生融合，形成了伸长的多核肌肉类似细胞，利用肌肉特异蛋白的抗体将这些细胞染成绿色。在背景中未表达 *myoD* 基因的成纤维细胞几乎看不到。（B）一些参与骨骼肌发育的基因调节蛋白的简图。中胚层祖细胞发育进入肌肉特异的途径涉及 4 种生肌基因调节蛋白的合成，它们分别是 MyoD、Myf5、肌细胞生成蛋白（成肌素）和 Mrf4。这些蛋白质直接激活肌肉结构基因以及 *MEF2* 基因的转录，后者编码另外一种基因调节蛋白。Mef2 与生肌蛋白联合发挥作用以便进一步激活肌肉结构基因的转录，并形成一个正反馈环，其作用是维持生肌基因的转录[（A）经 Stephen Tapscott and Harold Weintraub 许可；（B）摘自 J. D. Molkentin and E. N. Olson, *Proc. Natl. Acad. Sci.* USA 93：9366～9373，1996]

真核生物中组合基因调控产生许多不同的细胞类型

我们已经讨论过多个基因调节蛋白如何协同作用调节单个基因的表达。但是，正如生肌蛋白所表明的那样，组合基因调控的意义不止这一点：不仅每个基因拥有许多基因调节蛋白来调控它，而且每个基因调节蛋白也参与调控多个基因。此外，虽然有些基因调节蛋白对单个细胞类型特异，但是大多数在多种细胞类型，体内多个部位和发育期间多次打开。图 7-73 概括地说明了这一点，它显示了组合式基因调控如何能够以相对较少的基因调节蛋白产生大量的生物学复杂性。

利用组合式调控，一个给定的基因调节蛋白作为一组特定基因的指挥官或者一个特定细胞类型的规定者，并不是必定有一个可以简单界定的功能。相反，基因调节蛋白可以比作某种语言的单词：它们在不同的上下文中用于不同的意思并且基本不会单独使用；正是精心选出的组合传递了规定基因调节事件的信息。组合式调控的一个条件是许多基因调节蛋白必须能共同作用来影响最终的转录速率。这个原理在极其显著的程度上是对的：如果用实验方法引进同一细胞中，即使来自差别巨大的真核物种的不相关基因调节蛋白也能协同作用。这种

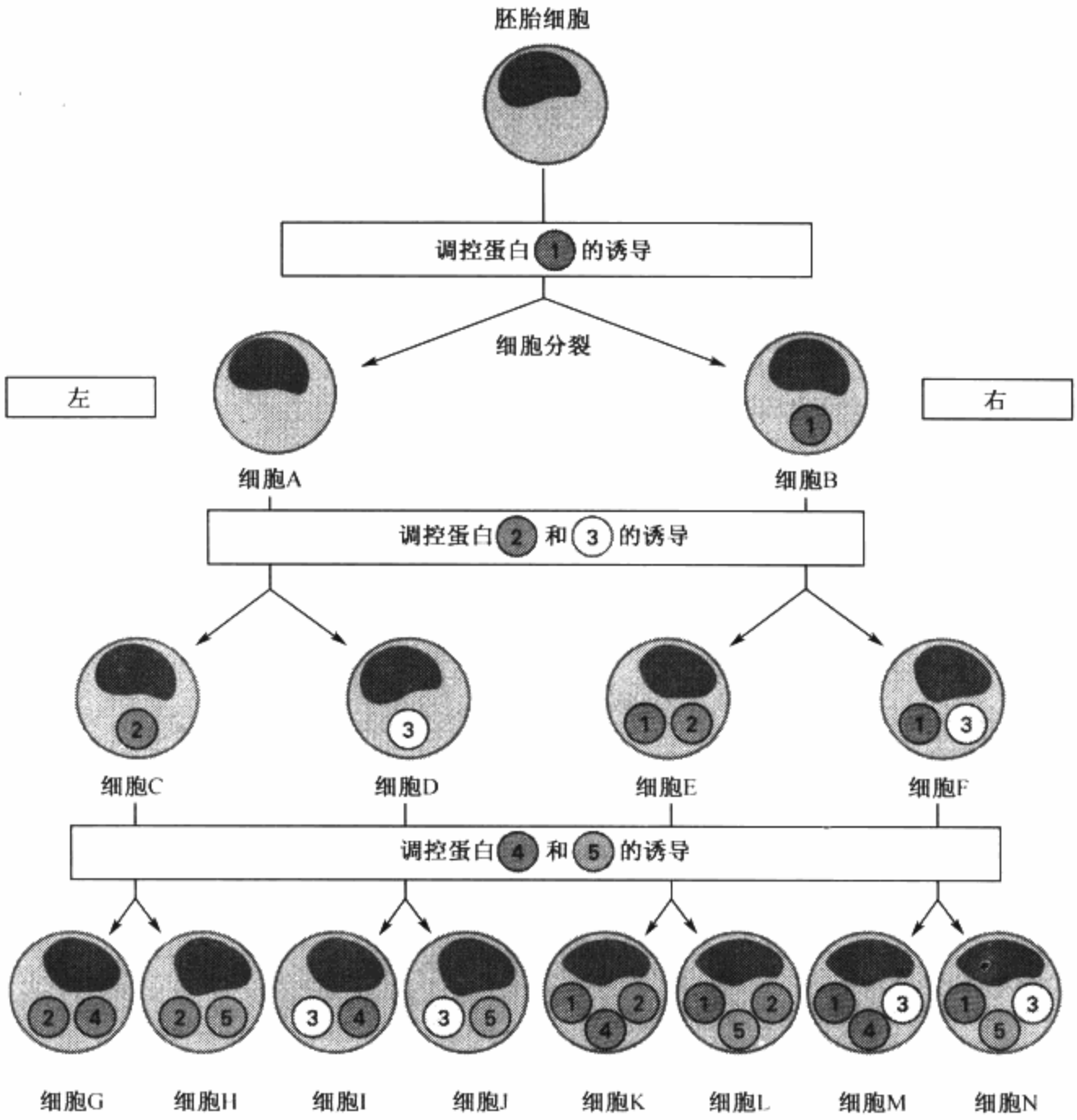


图 7-73　组合基因调控对发育的重要性

发育过程中一些基因调节蛋白的组合能产生许多细胞类型。在这个简单并且理想化的体系中，每一次细胞分裂之后就会做出一个决定，合成一对不同基因调节蛋白的其中一个（用标上数字的圆圈表示）。感受到在胚胎中的相对位置以后，朝向胚胎左侧的子细胞常常诱导合成每对蛋白质中的偶数蛋白质，而朝向胚胎右侧的子细胞诱导合成奇数蛋白质。假设每种基因调节蛋白的合成一旦起始以后就自我持续下去（图 7-68）。以这种方式，通过细胞记忆，逐步建立最终的组合指令。在此纯粹假设的例子中，利用 5 种不同的基因调节蛋白最终形成 8 种细胞类型（G～N）

情况既反映了转录装置高度保守性，也反映了转录活性自身的特性。我们已经看到，多种激活蛋白在转录协同中对转录的最终状态能表现出比加和效应更多的影响，而转录协同决定于转录装置应答多种输入信号的能力（图 7-47）。看起来似乎基因调节蛋白作用的多功能组合模式对它们的进化产生了严格的限制：它们必须与其他的基因调节蛋白、通用转录因子、RNA 聚合酶全酶和染色质修饰酶相互作用。

基因组合式调控的一个重要结果就是，向细胞中增加一个新基因调节蛋白的影响将决定于细胞过去的历史，因为这个历史将决定哪个基因调节蛋白已经存在。因此在发育过程中，细胞累积许多基因调节蛋白，它们起初不需要改变基因表达。然而，当加入基因调节蛋白必

需组合中的最后一个成员时，调节信息就终止了，导致基因表达发生巨大的变化。这种策略，我们已经看到，有助于解释向成纤维细胞加入单个调节蛋白质时如何能产生显著的转变，使成纤维细胞成为肌细胞。它也能解释第 21 章讨论的细胞决定过程和细胞分化过程之间的重要差别——前者中细胞定型为走向特定的发育命运，后者中已定型的细胞表达其特化的特征。

单个基因调节蛋白能触发整个器官的形成

我们已经看到，即使组合式调控是真核基因的标准调控方式，但是单个基因调节蛋白如果完成了适当组合的组装，就能在打开或关闭一整套基因中起决定性作用，并且我们已经看到了这种方式如何将一种细胞类型转变成另一种细胞类型。这个原理的一个显著的延伸来自于对果蝇、小鼠和人类眼睛发育的研究。这里有一个基因调节蛋白（在果蝇中称为 Ey，在脊椎动物中称为 Pax-6）很关键。如果在适当的情况下表达，Ey 能触发形成的不只是一种细胞类型，而是整个器官（一只眼睛），它由不同类型的细胞形成，并全部在三维空间中正确组织起来。

有关 Ey 作用最显著的证据来自于在果蝇中进行的实验，将 *ey* 基因人为地在发育早期通常形成腿部的一组细胞中进行表达。这种异常的基因表达导致在腿的中央长出眼睛（图 7-74）。果蝇的眼睛由数千个细胞组成，一个调节蛋白如何协同组织中全部的指令是发育生物学的中心主题。第 21 章将会讨论，它涉及细胞-细胞相互作用，以及细胞内基因调节蛋白。这里，我们注意 Ey 通过结合许多其他基因的调控区域，直接调控了它们的表达。这些受 Ey 调控的基因有些本身也是基因调节蛋白，它们继而再调控其他基因的表达。此外，这些基因中有些反过来作用于 *ey* 自身，从而形成一个正反馈环，确保了 Ey 蛋白的持续合成（图 7-75）。通过这种方式，仅仅一种调节蛋白的作用就能打开大量基因调节蛋白和细胞-细胞相互作用机制，该机制的作用导致形成许多组织有序的不同类型的细胞。可以想像到一个复杂的生物通过反复应用这一原理，是怎样一小部分一小部分地组建起来。

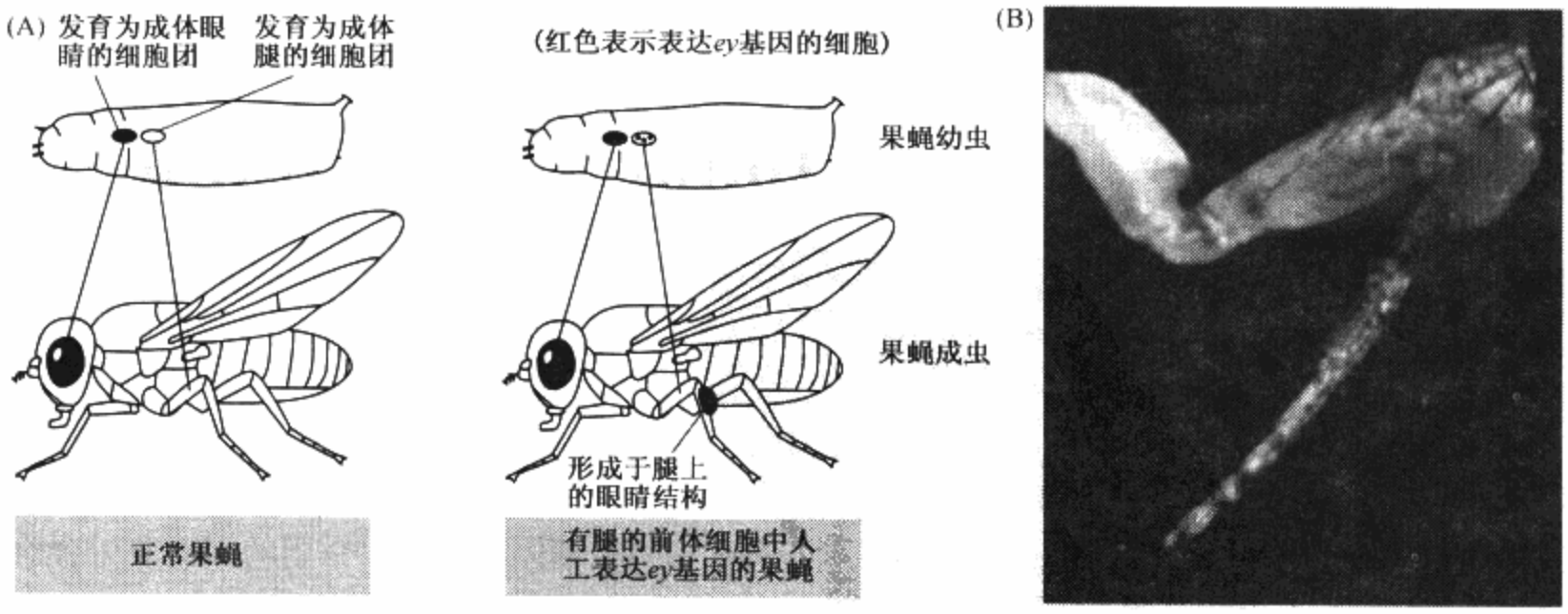


图 7-74　在腿的前体细胞中表达果蝇 *ey* 基因导致在腿上长出眼

(A)当果蝇幼虫或者含有正常表达的 *ey* 基因(左边)，或者含有在通常发育成腿部组织的细胞内人为地额外表达的 *ey* 基因(右边)时，显示其结果的简图。(B)一条畸形腿的照片，其上含有一只错位的眼[(B)经 Walter Gehring 许可]

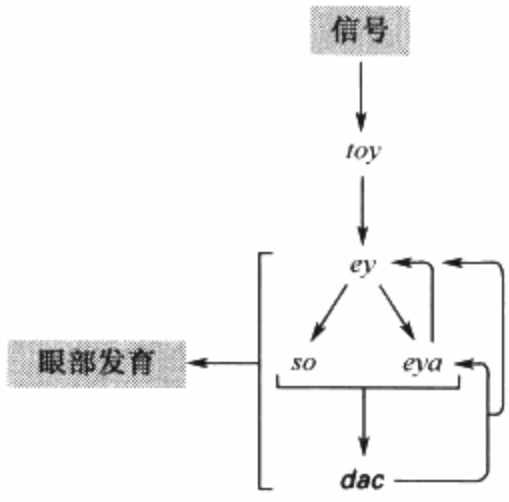


图 7-75　果蝇体内指令眼部发育的基因调节蛋白

toy(twine of eyeless)和 *ey*(eyeless)编码类似的基因调节蛋白，Toy 和 Ey，它们每个异位表达时都能引发眼部的发育。在正常的眼部发育过程中，*ey* 的表达需要 *toy* 基因。一旦 Toy 激活了 Ey 的转录，Ey 就激活 *so*(sine oculis)和 *eya*(eyes absent)的转录，这两个产物共同作用表达 *dac*(dachshund) 基因。如绿色箭头所示，有些基因调节蛋白形成正反馈环，它加强了起初向眼部发育的定型。已知 Ey 蛋白直接结合参与眼部发育的大量目标基因，包括编码晶状体晶体蛋白（图 7-119）、视紫红质和其他光受体蛋白的基因(摘自 T. Czerny et al. *Mol. Cell* 3:297～307,1999)

稳定的基因表达模式能够传递到子代细胞

一旦生物体中的一个细胞分化成特定的细胞类型，它一般就以那种方式保持分化状态，如果它发生分裂，它的子细胞就会继承相同的分化特征。例如，肝细胞、色素细胞和内皮细胞（讨论见第 22 章）在个体生命中分裂许多次。这就意味着一个分化细胞特异的基因表达模式必须被记忆并经由随后进行的细胞分裂传递到后代细胞中。

我们已经讨论了多个确保子细胞能“记住”它们最终将成为的细胞类型的方式。最简单的一种是通过正反馈环[图7-68、图7-72(B)和图7-75]。这里有一个重要的基因调节蛋白激活它自身基因以及其他细胞类型特异基因的转录（直接地或间接地）。图 7-67 中显示的简单转变开关是这种规则的一种变化形式：通过抑制其自身阻遏蛋白的表达，一个基因产物间接地激活并维持它自身的表达。真核生物中另外一个非常不同的维持细胞类型的方式是通过染色质结构从亲代细胞向子代的忠实传递，这些已在第 4 章讨论。一旦基因调节蛋白规定了一种分化细胞类型，就可以通过将未表达的基因包装成更紧密的染色质，并将该染色质做上“沉默”标记来加强发育决定（图 4-35）。活跃转录的基因也能以相同的机制作上标记和进行传递。基因组选定区域包装成凝聚的染色质是一种细菌所不具备的基因调控机制，据认为它使真核细胞经历很多世代以后仍然特别稳定地维持基因表达模式。这种稳定性对于多细胞生物特别关键，因为在多细胞生物中单个细胞中异常表达的基因能对整个生物体发育产生影响深远的后果。

如果基因表达模式的维持决定于染色质包装模式，那么这种染色质构型如何从一个细胞传递到它的细胞中呢？第 4 章已经讨论了一些可能的机制（图 4-48）。有一个通用的机制依赖于蛋白质对 DNA 的协同性结合（图 7-76）。当细胞复制 DNA 时，每条 DNA 链都能继承一份结合在原来 DNA 双螺旋特定区段上的蛋白质分子，这些继承得到的分子接着能招募新合成的分子，在每个子细胞中重建出起初的染色质复合体的一份完整拷贝。这种细胞记忆机制可以基于特殊基因调节蛋白的协同性结合，或者是普通染色质结构组分的协同性结合，或者这两类分子共同作用。因此，特殊基因调节蛋白最初的某种结合模式能起始形成一种染色质凝

聚模式，并随后将这种模式维持起来。

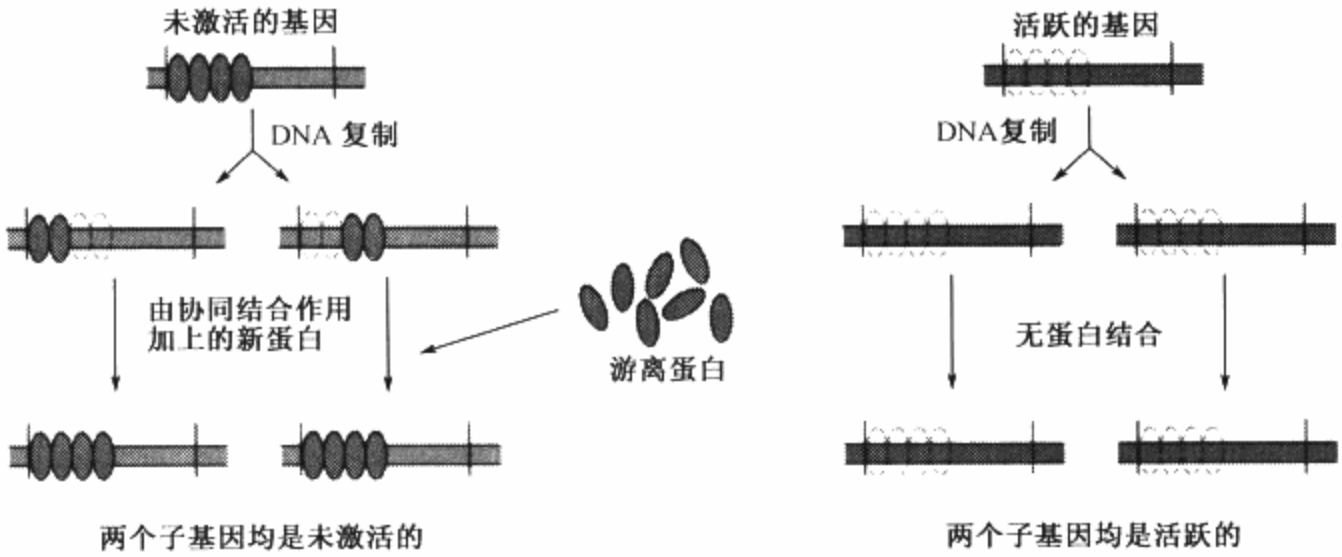


图 7-76　DNA 复制期间基因表达状态得以直接遗传的一般体系

在此假设的模型中，协同结合的染色体蛋白质簇的一部分直接从亲代 DNA 双螺旋（左上）转移到两个子代螺旋中。遗传得到的蛋白质簇然后引起每个子代 DNA 螺旋结合相同蛋白质的其他拷贝。因为结合是协同的，所以从缺少结合蛋白的相同亲代 DNA 螺旋（右上）合成得到的 DNA 上仍然无结合蛋白。如果结合蛋白关闭基因转录，那么无活性的基因状态将直接遗传，如图所示。如果结合蛋白激活了转录，那么活跃的基因状态就将直接遗传（未显示）。当协同的蛋白结合需要特殊 DNA 序列时，这些事件将限制在特定基因调控区发生；然而，如果结合能沿染色体扩展，那么就可以解释与第 4 章讨论的可遗传染色质状态关联的扩散效应。

虽然蛋白质都描绘成一样的，但相同的原理也能解释不同蛋白质组合协同组装如何能够稳定地扩散

然而细胞记忆采取的另一种策略基于染色质蛋白酶催化修饰的自我增殖模式（我们在第 4 章看到过），或者甚至是 DNA 自身，这一点我们将在后面解释。但是首先我们更仔细地看一个特殊的例子，其中细胞记忆明显地涉及染色质结构的改变。

染色质结构中广泛的染色体改变可以遗传

第 4 章我们看到染色质状态是可以遗传的，并且它们可以用于建立和保存跨过 DNA 很长距离和经过许多细胞世代的基因表达模式。这种染色质组织长程效应的一个显著例子存在于哺乳动物中，在这里整条染色体的染色质结构变化用于调节这条染色体上全部基因的表达。

雄性和雌性的性染色体不同。雌性有两条 X 染色体，而雄性有一条 X 和一条 Y 染色体。结果雌性细胞所含 X 染色体拷贝数目是雄性细胞的 2 倍。在哺乳动物中，X 和 Y 染色体的基因含量存在根本性的差别：X 染色体很大，含有 1000 多个基因，而 Y 染色体较小含有不到 100 个基因。哺乳动物演化出一种剂量补偿机制来平衡雄性和雌性之间 X 染色体基因产物的数量。干扰剂量补偿的突变是致死的，从而证明必须维持 X 染色体与常染色体（非性染色体）基因产物的比率。

在哺乳动物中通过失活雌性体细胞中的一条 X 染色体的转录来实现剂量补偿，这个过程称为 **X 染色体失活**。在雌性胚胎发育早期，此时它由几千个细胞组成，每个细胞中的两条 X 染色体会有一条高度凝聚成一种异染色质。间期细胞中凝聚的 X 染色体能在光学显微镜下容易地观察到；它最初被称为巴氏小体，位于核膜附近。作为 X 染色体失活的结果，两条 X 染色体能够共存于接触着相同转录调节蛋白的同一细胞核中，但是它们的表达情况完全不同。

最初失活哪一条 X 染色体的选择，是从母代细胞继承的那条（Xm）还是从父代细胞继承的那条（Xp），是随机进行的。一旦 Xm 或 Xp 已经失活，那么在该细胞及其后代随后进行

的所有细胞分裂中，它都保持沉默，这表明经过许多轮 DNA 复制和减数分裂，仍然忠实地维持着失活状态。由于 X 染色体失活是随机的，并在已经有几千个细胞形成胚胎之后才进行，所以每个雌性个体是沉默了或者 Xp 或者 Xm 的细胞克隆群组成的嵌合体（图 7-77）。这些克隆群在成年动物中成小串分布，因为姐妹细胞倾向于在发育后期保持紧靠在一起。例如，在有些雌猫中，X 染色体失活导致产生红色和黑色“龟壳”被毛颜色。在这些猫的体内，一条 X 染色体携带一个产生红色毛色的基因，另一个 X 染色体携带同一基因的等位基因，导致产生黑色毛色；正是随机的 X 染色体失活导致形成两种不同颜色区域的细胞。与雌性不同，这种基因品种的雄猫要么是全红要么是全黑，全由它从母亲继承得到的是哪条 X 染色体决定。虽然经过数千次细胞分裂以后仍然维持着 X 染色体失活，但并不是永久性的。尤其是在形成生殖细胞时会发生逆转，这样所有的单倍体卵母细胞含有一条活跃的 X 染色体，并能够表达 X 染色体相关的基因产物。

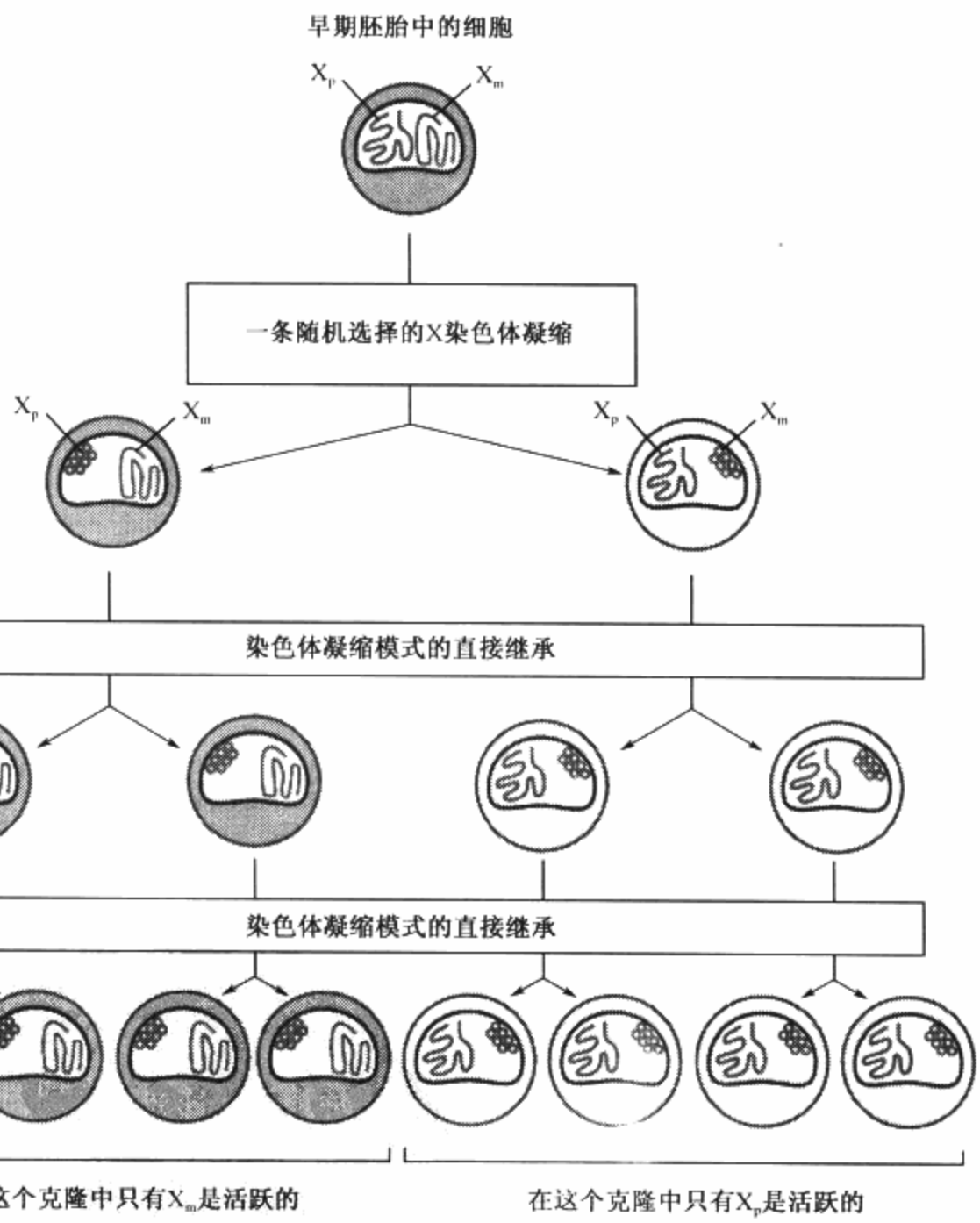


图 7-77　X 染色体失活

雌性哺乳动物中凝缩的失活 X 染色体的克隆遗传

整条染色体是如何失活的呢？X 染色体失活起始于单个位点并从此处扩展开，这个位点位于 X 染色体中央，称为 **X 染色体失活中心（XIC）**。从 XIC 取出的一部分 X 染色体与常染

色体融合，就不会失活。相比之下，与失活 X 染色体的 XIC 融合的常染色体发生转录沉默。因而可以将 XIC（一段将近 10^6 核苷酸对的 DNA 序列）看作是一个巨大的调节元件，它是异染色质形成的起点，并促进它沿整条染色体双向延伸。XIC 中编码了一种不常见的 RNA 分子，XIST RNA，它只在失活的 X 染色体中表达，并且其表达对于 X 染色体失活是必需的。XIST RNA 并不翻译成蛋白质，而是保留在核内，最终将失活的 X 染色体包被起来。XIST RNA 从 XIC 跨越整条染色体进行延伸与基因沉默的扩展有关，表明 XIST RNA 参与了异染色质的形成和延伸（图 7-78）。除含有 XIST RNA 以外，X 染色体的异染色质还表现出其他特征，包括拥有组蛋白 2A 的一种特殊变体，组蛋白 H3 和 H4 的乙酰化水平低，组蛋白 H3 一个特定位点上发生甲基化，以及所含 DNA 发生甲基化，这个话题我们将在下面讨论。有可能所有这些特征使得失活的 X 染色体常常能抵制转录。

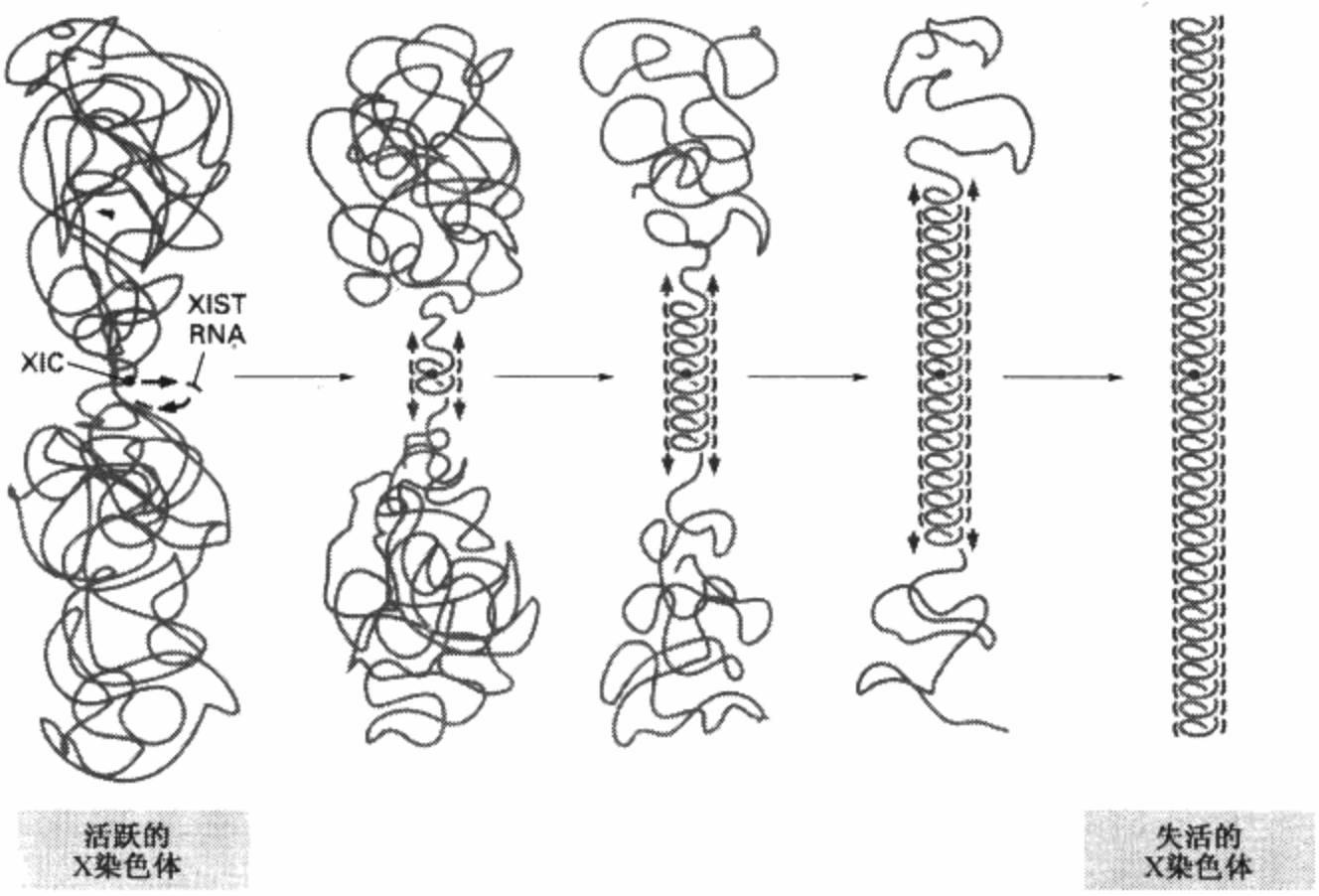


图 7-78　哺乳动物的 X 染色体失活

X 染色体失活开始于从 XIC（X 染色体失活中心）基因座合成 XIST（X 染色体失活特异转录物）。XIST RNA 与 X 染色体结合同染色体凝缩有关。XIST 的结合与染色体凝缩都从 XIC 基因座逐渐向外朝染色体末端移动。有关此过程如何进行的细节还有待研究

哺乳动物 X 染色体的许多特征仍然有待发现。最初是如何决定将哪条 X 染色体失活的？又是什么机制防止了其他 X 染色体也被失活？XIST RNA 是怎样协调异染色质形成的？失活的染色体经过很多次细胞分裂后又是怎样维持的？我们才刚刚了解对于我们人类自身这个物种的生存很关键的这个基因调控机制。

雌性个体中 X 染色体失活是有性生殖生物解决剂量补偿问题的唯一方式。在果蝇中，雄性细胞中的单条 X 染色体上全部基因的转录水平比雌性细胞中单条 X 染色体的基因高 2 倍。雄性 X 染色体的染色质结构改变导致雄性特异的转录“上调”。对于在哺乳动物中，这种改变涉及一种特殊 RNA 分子与 X 染色体的结合；然而在果蝇中，与 X 染色体结合的 RNA 增加了基因活性而不是阻遏了它的活性。雄性 X 染色体也具有组蛋白乙酰化的特殊模式，它可能有助于吸引转录装置到这条染色体上来（图 4-35 和图 7-46）。

在线虫中通过另一种策略来实现剂量补偿。这里两种性别分别是雄性（有一条 X 染色体）和雌性同体（有两条 X 染色体）。通过雌雄同体细胞中两条 X 染色体每条转录都“下调”1 倍来实现剂量补偿。这一点的实现是由于雌雄同体的 X 染色体整体结构发生了大范围改变（图 7-79）。这些变化涉及 X 染色体特异的蛋白质组装，其中有些是共有的凝聚蛋白，它们在有丝分裂过程中帮助凝聚染色体（图 4-56 和图 18-3）。

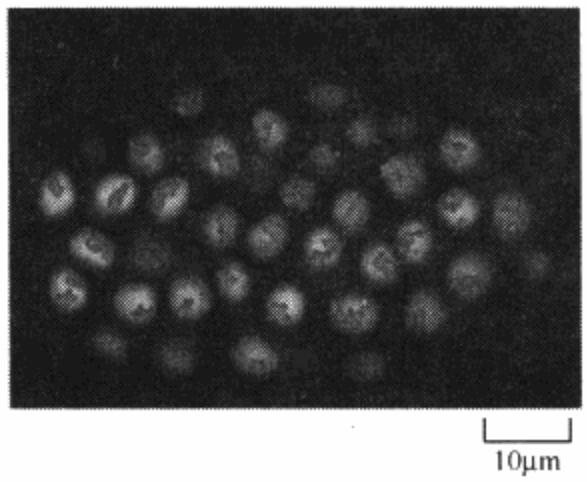


图 7-79　剂量补偿蛋白在线虫雌雄同体（XX）细胞核的 X 染色体上的定位

图片中可看到许多来自发育胚胎的细胞核。用 DNA 嵌入染料 DAPI 将全部 DNA 染成蓝色，用偶联上荧光染料的抗 Sdc-2 抗体将 Sdc-2 蛋白染成红色。该实验显示 Sdc-2 蛋白只与有限的一组染色体结合，经其他实验鉴定后是两条 X 染色体。Sdc-2 沿全长的 X 染色体结合，并吸引其他蛋白质，包括凝聚蛋白类似复合物，形成这些染色体完整的特殊结构（摘自 H. E. Dawes et al., *Science* 284：1800～1804，1999. © AAAS）

虽然哺乳动物、果蝇和蠕虫实现剂量补偿的策略不同，但它们都涉及整个 X 染色体结构的改变。有可能在进化中，为了克服有性生殖动物在基因调控中遇到的高度特异性问题，改变并利用了染色体结构的一般特征。

DNA 甲基化模式在脊椎动物细胞分裂时可以遗传

到此为止，我们已经着重讨论了蛋白质与特异 DNA 序列联合进行的基因转录调控。然而 DNA 自身可以发生共价修饰，在下面的部分我们将看到这种修饰也为基因表达提供了调节机会。在脊椎动物细胞中胞嘧啶甲基化似乎提供了一种区分活性基因和非活性基因的重要机制。胞嘧啶的甲基化形式，5-甲基胞嘧啶（5-甲基 C）和胞嘧啶的关系与胸腺嘧啶和尿嘧啶的关系相同，并且这种修饰同样对碱基配对没有影响（图 7-80）。脊椎动物 DNA 中的甲基化限制在 CG 序列的胞嘧啶（C）核苷酸上，这个序列恰恰与 DNA 双螺旋另一条链上的 CG 序列配对（方向相反）。因此，存在一种简单的机制使现有的 **DNA 甲基化**模式直接由子代 DNA 链继承。一种维持甲基化转移酶优先作用于和甲基化的 CG 配对的 CG 序列上。结果，亲代 DNA 链上 DNA 甲基化模式为子代 DNA 链甲基化提供了模板，导致这种模式随 DNA 复制后直接遗传下去（图 7-81）。

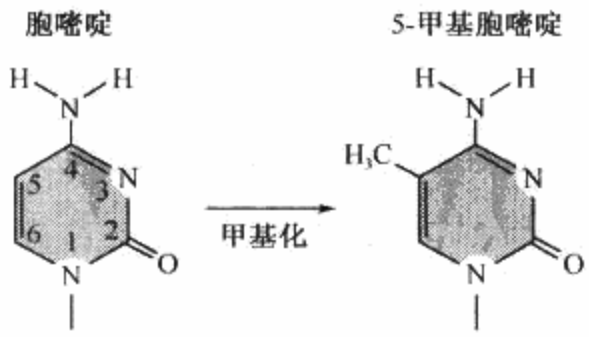


图 7-80　DNA 双螺旋中胞嘧啶碱基甲基化形成 5-甲基胞嘧啶

在脊椎动物中该事件只限于在经过选择的 CG 序列中的胞嘧啶核苷酸上发生

DNA 甲基化模式的稳定遗传可以由维持 DNA 甲基化转移酶解释。然而 DNA 甲基化模式在脊椎动物的发育期间是动态变化的。在受精后不久，会发生整

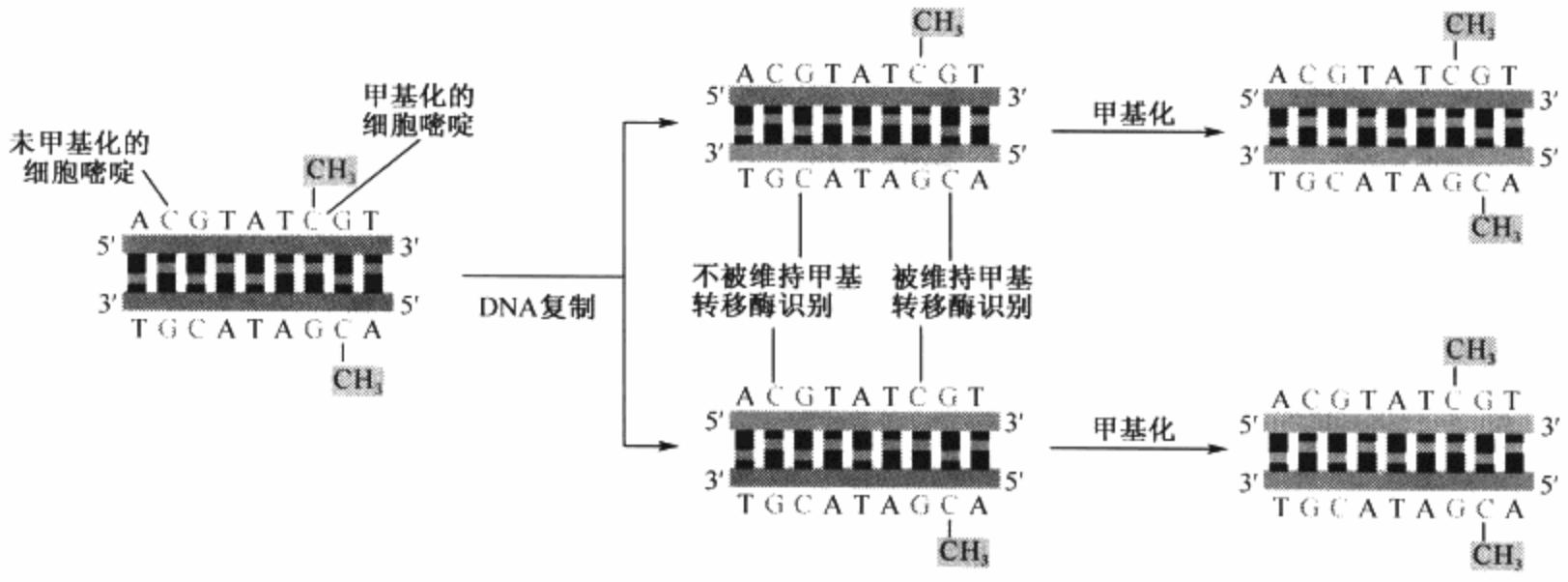


图 7-81　DNA 甲基化模式如何忠实地遗传

脊椎动物 DNA 中 CG 序列上大部分胞嘧啶核苷酸都甲基化（图 7-80），由于甲基指导的甲基化酶（维持甲基转移酶）的存在，一旦确立了一种 DNA 甲基化模式，每个甲基化位点就会在后代 DNA 中遗传，如图所示

个基因组范围的去甲基化，此时大多数甲基从 DNA 丢失。这种去甲基化的进行可以是通过抑制维持 DNA 甲基转移酶的活性，从而每轮 DNA 复制期间甲基发生被动的丢失，或者是通过特殊的去甲基化酶的作用。在发育晚期，当胚胎植入子宫壁时，由多个修饰特定未甲基化 CG 双核苷酸的从头 DNA 甲基转移酶建立新的甲基化模式。一旦建立了新甲基化模式，它们就能经过多轮 DNA 复制由维持甲基转移酶进行传递。维持或者从头甲基转移酶发生突变都将导致小鼠的早期胚胎死亡，表明建立和维持正确的甲基化对于正常发育是关键的。

脊椎动物利用 DNA 甲基化将基因锁定在沉默状态

在脊椎动物中，发现 DNA 甲基化主要位于基因组转录沉默区域。例如，失活的 X 染色体或某些组织中失活的基因，从而表明它在基因沉默中起作用。脊椎动物细胞含有一个结合甲基化 DNA 的蛋白质家族。这些 DNA 结合蛋白质继而与染色质重建复合体和组蛋白脱乙酰化酶相互作用，它们凝聚染色质，使染色质转录失活。即便如此，DNA 甲基化不足以提供基因失活的信号，下面的例子将证明这些。在体外利用去甲基化和甲基化 DNA 的细菌蛋白，可以制备完全甲基化和完全去甲基化的编码肌肉特异的肌动蛋白基因的质粒 DNA。当这样的两个质粒引入培养的肌肉细胞后，甲基化的质粒与未甲基化的质粒会以相同的高速率转录。而且在正常发育过程中，如果一个沉默，甲基化的基因被打开，只有在该基因转录一段时间以后才发生甲基化丢失。最后，在 X 染色体失活期间，凝缩和沉默在 DNA 甲基化水平增加以前进行。这些结果都表明甲基化加强了起初由其他机制建立的转录阻遏。DNA 甲基化似乎在脊椎动物中主要用于确保一旦基因关闭以后，它能完全保持关闭状态（图 7-82）。

为了检验在一种脊椎动物细胞类型中高水平转录的一段 DNA 序列是否在另一种细胞类型中无论如何都会转录，所设计的实验已经表明在两种细胞类型中基因转录速率相差可超过 10^6 倍。因此就转录而言，脊椎动物中未表达的基因比细菌中未表达的基因更少会发生“渗漏”，已知在细菌中表达和未表达基因状态之间最大的转录速率差别大约是 1000 倍。脊椎动物未表达基因的 DNA 甲基化以及因之产生的染色质结构变化，至少导致产生了这些差别中的一部分。每个脊椎动物细胞中正常情况下完全关闭的成千上万的基因发生“渗漏”转录可

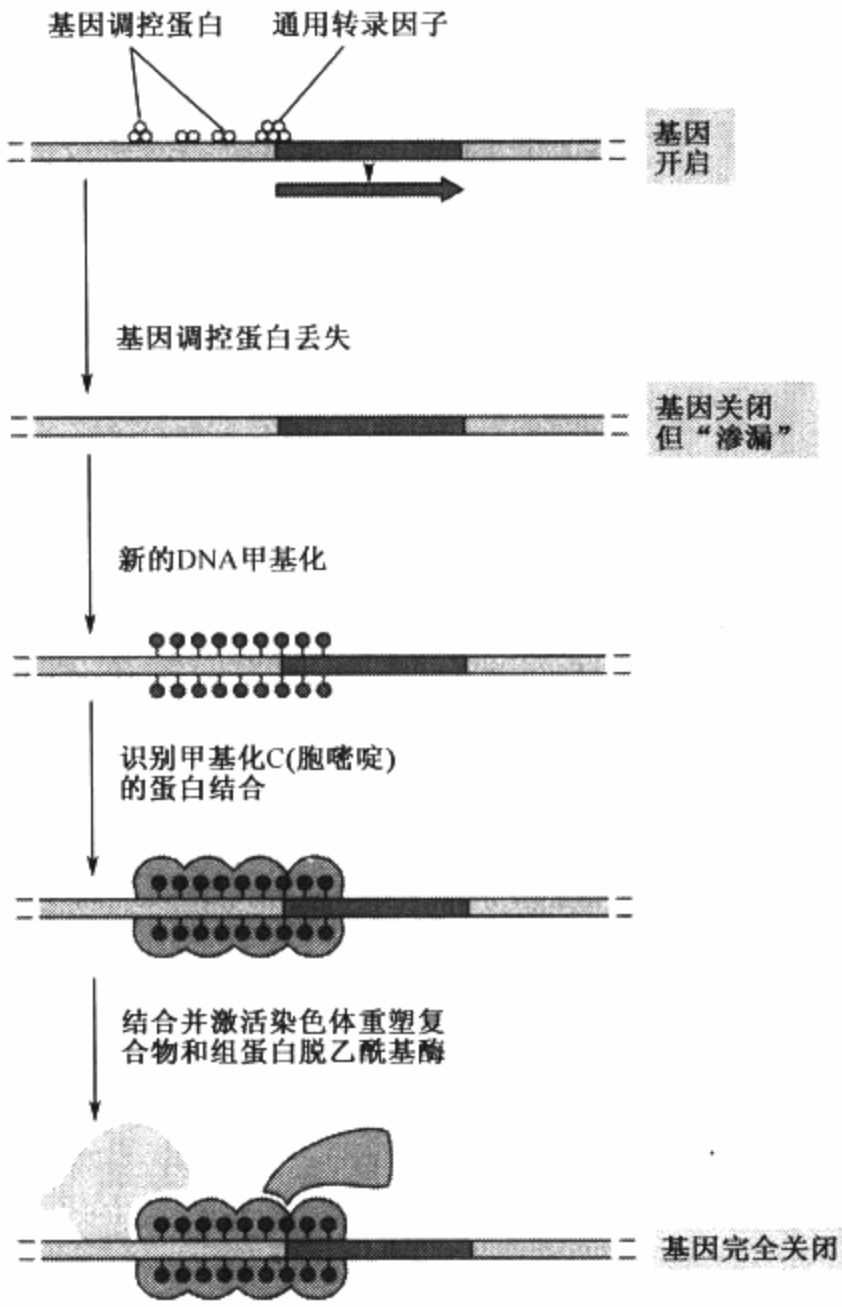


图 7-82　DNA 甲基化如何关闭基因

基因调节蛋白和通用转录装置在活性启动子附近的结合，可能通过排除从头甲基化酶阻止了 DNA 的甲基化。然而，如果大多数这些蛋白质与 DNA 解离，正如细胞不再合成所需的激活蛋白时通常会发生的那样，DNA 就发生甲基化，使其他蛋白质能够结合，这些蛋白质通过进一步改变染色质结构完全关闭基因（图 7-49）

能是缺乏维持 DNA 甲基转移酶的小鼠中早期胚胎死亡的原因。

转录沉默在脊椎动物基因组中对于抑制转座元件的扩散也特别重要（图 4-17）。尽管编码序列只占典型脊椎动物基因组的几个百分点，但转座元件几乎能占这些基因组的一半。我们在第 5 章看到过，转座元件能够自我拷贝并将这些拷贝插入基因组其他位置，就有可能打断基因或者重要的调节序列。通过阻止转座元件的转录，DNA 甲基化限制了它们的散播，因而维持了基因组的完整性。除了这些不同的用处之外，DNA 甲基化也至少是一类特殊类型的细胞记忆所必需的，我们下面接着讨论。

基因组印迹需要 DNA 甲基化

哺乳动物细胞是二倍体，含有分别从父亲和母亲遗传下来的一套基因。在少数情况下发现一个基因的表达决定于它是从父亲还是从母亲遗传得到，这个现象称为**基因组印记**。胰岛素类似生长因子-2 的基因（*Igf2*）就是印记基因的一个例子。*Igf2* 为出生前的生长所需，

不表达 *Igf2* 的小鼠出生时大小是正常小鼠的一半。只有父源的 *Igf2* 的小鼠出生时大小是正常小鼠的一半。只有父源的 *Igf2* 拷贝发生了转录，结果父源 *Igf2* 基因有一个突变的小鼠就变得短小，而母源 *Igf2* 有缺陷的小鼠却是正常的。

在生殖细胞形成的过程中，发生印记的基因由甲基化根据它们是存在于精子还是卵细胞中进行标记。通过这种方式父源基因随后可以在胚胎中识别；因此甲基化被用作标记来区分一个基因的两个拷贝，而它们在其他情况下是相同的（图 7-83）。由于受精后不久发生的去甲

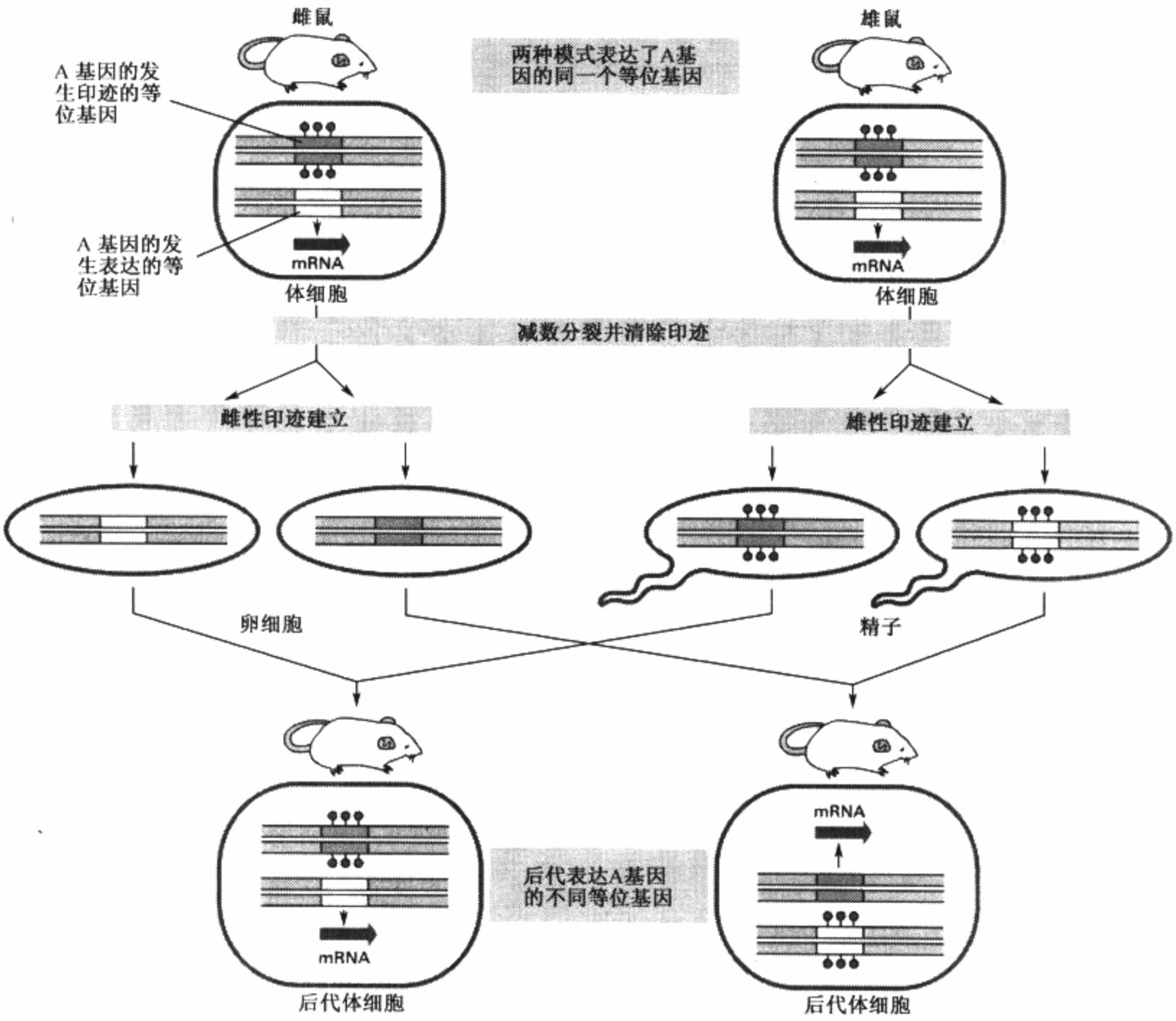


图 7-83　小鼠中的印记

图的上部分显示了位于体细胞中的一对同源染色体，分别来自两只成年小鼠，一只雄鼠和一只雌鼠。本例中，两只鼠都从它们的父亲继承了上面的同系物，从母亲继承了下面的同系物。发生印记（用橙色表示）的父本拷贝甲基化，这样就阻止它的表达。相同基因（黄色）的母源拷贝则发生表达。图的剩余部分显示了这两只鼠杂交的结果。在减数分裂和形成生殖细胞的过程中，起初消除印记，然后重新恢复（图的中间部分）。在雌鼠产生的卵子中，A 基因任何一个等位基因都不甲基化。在雄鼠产生的精子中，A 基因两个等位基因都甲基化。图的底部显示的是后代鼠可能继承的两种印记模式；左边的小鼠具有与父母任何一个都相同的印记模式，而右边的小鼠则具有相反的模式。如果 A 基因的两个等位基因不同，这些不同的印记模式能引起后代小鼠中表型的差别，即便它们确实携带了 A 基因这两个等位基因的相同 DNA 序列。印记是经典遗传行为的一个重要例外情况，据认为有 100 多个小鼠基因以这种方式受到影响。然而大多数小鼠基因未发生印记，因而孟德尔遗传定律适用于小鼠基因组的大部分

基化波不影响印记基因，所以这个标记使体细胞能“记住”该基因两个拷贝中的父源拷贝，并据此调节它们的表达。在大多数情况下，甲基化印记利用图 7-82 所示的机制使周围的基因发生沉默。然而有些情况下，甲基化印记能激活基因的表达。对于 *Igf2* 而言，例如，父源染色体上一个绝缘子元件（图 7-61）发生甲基化而阻断其功能，并使远处的增强子能够激活 *Igf2* 基因的转录。在母源染色体上，绝缘子不被甲基化，因而 *Igf2* 不发生转录（图 7-84）。

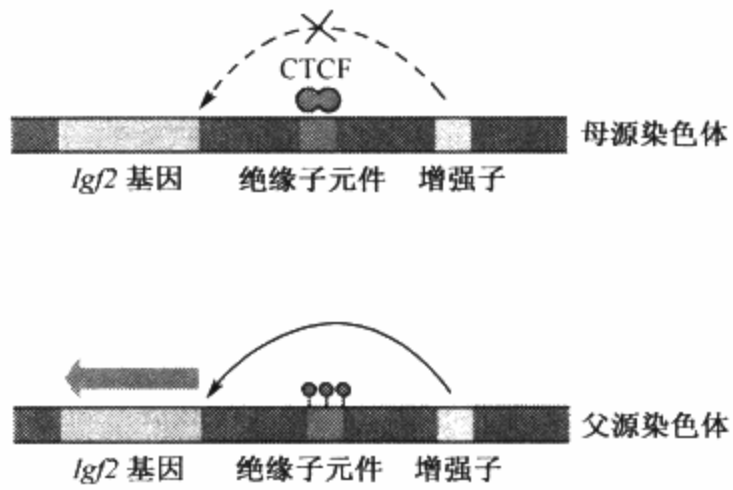


图 7-84　小鼠 *Igf2* 基因印记机制

在从雌鼠继承得到的染色体上，一种称作 CTCF 的蛋白质结合一个绝缘子，（图 7-61）从而阻断了增强子（绿色）和 *Igf2* 基因（橙色）之间的通讯。因而 *Igf2* 在从母代遗传的染色体中不表达。因为存在印记，从雄鼠继承的染色体上的绝缘子甲基化，这样通过阻断 CTCF 蛋白的结合而使绝缘子失活，并允许增强子激活 *Igf2* 基因的转录。在其他印记的例子中，甲基化通过干扰基因转录所需蛋白质的结合来阻断基因表达。受精后由合子继承的染色体上甲基化模式（印记）在后代中通过维持甲基转移酶进行维护（图 7-81）

印记是后生遗传的一个例子，也就是指非 DNA 核苷酸序列改变产生的可遗传的表型改变。为什么会存在印记还是一个谜。在脊椎动物中，它只限于胎盘哺乳动物，并且所有的印记基因都参与胎儿发育。一种观点认为印记反映了在进化竞争中产生更大后代的雄性和限制后代大小的雌性之间的妥协。不论印迹的目的可能是什么，它提供了一个令人吃惊的证据，表明 DNA 除核苷酸序列以外的其他特征也可以遗传。

哺乳动物中 CG 岛大约与 20 000 个基因相连

由于 DNA 修复酶发挥功能的方式，基因组中甲基化的胞嘧啶核苷酸在进化过程中倾向于被清除掉。未甲基化的胞嘧啶发生偶然的脱氨基会生成尿嘧啶，它并非 DNA 中正常存在的核苷酸，因而由 DNA 修复酶、尿嘧啶 DNA 糖苷酶识别。但是 5-甲基胞嘧啶偶尔脱氨基不能以这种方式得到修复，因为脱氨基产物是胸腺嘧啶，这样就与 DNA 中其他非突变产生的

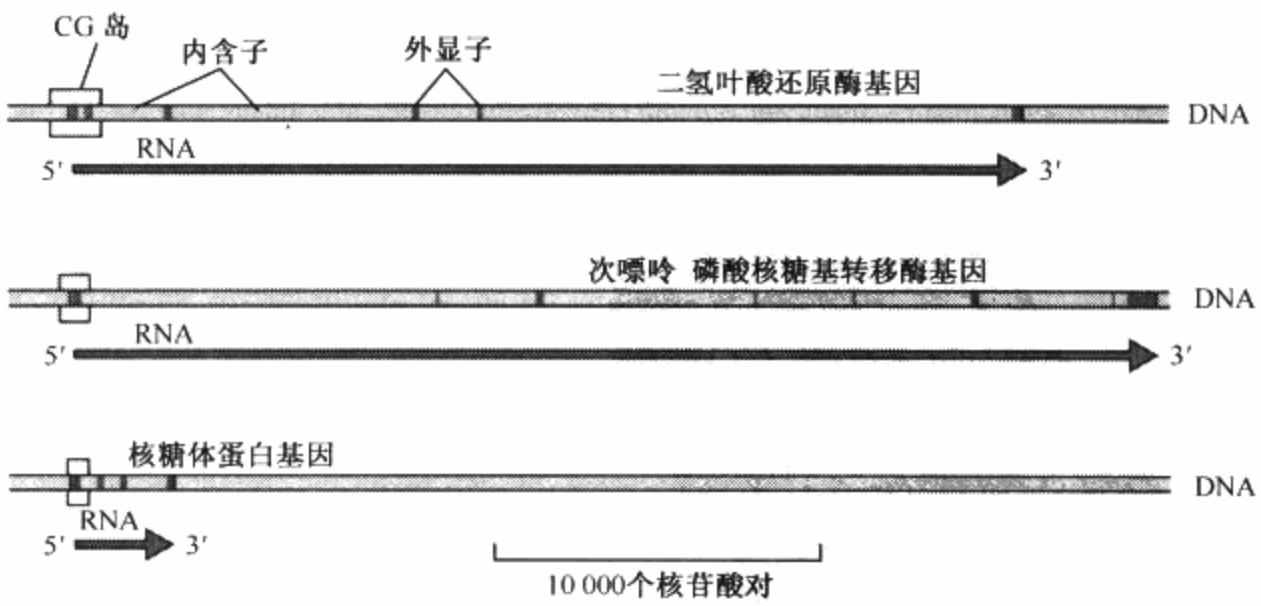


图 7-85　3 种哺乳动物持家基因中围绕着启动子的 CG 岛

黄色框表示每个 CG 岛的范围。对于哺乳动物中大多数基因来说（图 6-25），外显子（暗红色）相对于内含子（浅红色）而言非常短（摘自 A. P. Bird，*Trends Genet*. 3：342～347，1987）

胸腺嘧啶核苷酸难以区分。尽管存在一个专门的修复系统来除去这些 T 核苷酸突变物，但是许多脱氨基产物没有被发现，所以在进化期间基因组中那些甲基化的 C 核苷酸倾向于突变成 T。

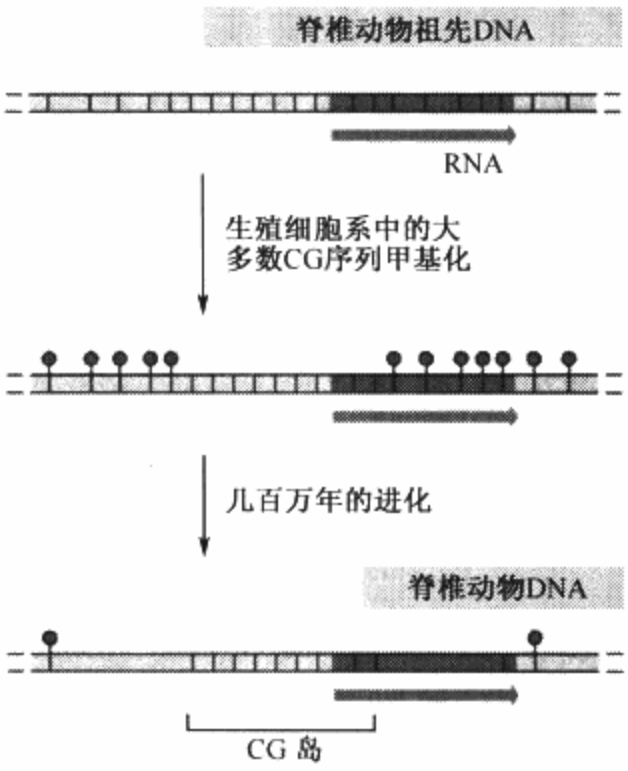


图 7-86　解释脊椎动物中整体上显著缺少 CG 序列以及它们簇集成 CG 岛的机制

黑线标记 DNA 序列中 CG 二核苷酸的位置，而红色“棒棒糖”表示 CG 二核苷酸上的甲基。位于生殖细胞转录基因的调节序列中的 CG 序列未甲基化，因而倾向于在进化中保留下来。另一方面，甲基化的 CG 序列倾向于通过 5 甲基胞嘧啶脱氨后变为胸腺嘧啶而丢失，除非该 CG 序列对生存很关键

在进化过程中，每 4 个 CG 中有超过 3 个以这种方式丢失掉，结果使脊椎动物显著缺少这种二核苷酸。保留下来的 CG 在基因组中分布非常不均匀；它们在长 1000～2000 个核苷酸对的称作 **CG 岛**的特定区域中存在密度是平均密度的 10～20 倍。这些 CG 岛似乎在所有细胞类型中都保持未甲基化。它们常常包围着所谓持家基因（指那些编码了对细胞存活极为重要的蛋白质因而在大多数细胞都表达的基因）的启动子（图 7-85）。此外，有些组织特异的基因（编码只在某些细胞类型中需要的蛋白质）也与 CG 岛关联在一起。

如果我们假设脊椎动物中采用 CG 甲基化主要是作为一种维持 DNA 处于转录失活状态（图 7-82 和图 7-86）的方式，那么 CG 岛（也称作 CpG 岛，以便区分 CG 二核苷酸和 CG 核苷酸对）的分布也可以得到解释。在脊椎动物中，只有当新的甲基化 C 向 T 的突变发生在产生精子或卵子的生殖细胞系时，才能传递到下一代。在脊椎动物生殖细胞中大多数 DNA 是失活的，且高度甲基化。在漫长的进化时间里，这些失活区域中甲基化的 CG 序列，可能通过自发的脱氨基事件且没有正确地加以修复而丢失掉。然而在生殖细胞系中活跃基因的启动子保持未甲基化，因而在它们中间发生 C 的自发脱氨基能得到精确地修复。这些区域在现今脊椎动物细胞中作为 CG 岛保留下来。此外，基因组中任何一个破坏了成年个体中某个基因的功能或调节的 CG 序列的突变，都将被选择，并且有些 CG 岛只是密度比关键 CG 序列正常的密度更高的结果。

哺乳动物基因组估计含有 20 000 个 CG 岛，大部分 CG 岛标记着转录单位的 5′端，所以也可能标记了基因的 5′端。CG 岛的存在常常提供了一种方便的方法，用于鉴定脊椎动物基因组 DNA 序列中的基因。

小结

动物和植物中许多细胞类型的产生，大都是通过使不同的基因在不同的细胞中表达的机制。由于许多特化的动物细胞能够经过许多次细胞分裂仍然维持它们独特的特征，甚至在培养基中生长的时候也如此，所以参与产生这些细胞的基因调节机制一旦确立后就必须保持稳定且在细胞分裂的时候可以遗传。这些特征赋予细胞记忆其发育历程的能力。细菌和酵母提供了一种非常容易利用的模型系统，用于在其中研究基因调节机制。这类机制之一涉及两个抑制彼此合成的基因调节蛋白之间竞争性的相互作用；这样可以形成一种转换开关，在两种可选择的基因表达模式之间改变细胞。直接或间接的正反馈环使基因调节蛋白能够永久地促

进它们自身合成，也提供了一种通用的细胞记忆机制。具有程序化延迟的反馈环构成了细胞钟的基础。

在真核生物中，基因的转录通常受到基因调节蛋白的组合来调控。有人认为高等生物中每个细胞类型都含有一种基因调节蛋白的特定组合，它确保了只有这些基因的表达对于这类细胞是合适的，特定的某个基因调节蛋白在不同环境中可能都是活跃的并且通常参与了许多基因的调控。

除了可扩散的基因调节蛋白以外，遗传得到的染色质凝聚状态也被真核生物用来调控基因表达。一个特别显著的例子就是雌性哺乳动物中整条 X 染色体的失活。在脊椎动物中 DNA 甲基化也在基因调控中发挥作用，主要用作一种机制来加强最初由其他机制做出的基因表达的决定。DNA 甲基化也是哺乳动物基因组印迹现象的原因，在此印迹中，一个基因的表达决定于它是来自母亲还是来自父亲。

转录后调控

从理论上讲，基因表达所需的每一步骤都能受到调控。事实上，人们可以找到每种调控类型的例子，尽管任何一个基因都只可能利用其中的一些。基因转录起始的调控是大多数基因的主要调控形式。但是其他的调控能够在从 DNA 到蛋白质途径的后阶段发挥作用，来调整所产生的基因产物数量。**转录后调控**在 RNA 聚合酶已经结合到基因启动子之后才发挥作用，虽然它们不如转录调控常见，但是对于许多基因来讲它们是很关键的。

在下面的部分我们根据一个 RNA 分子在转录起始以后可能经历的事件的顺序，按照时间先后讨论各种转录后调控（图 7-87）。

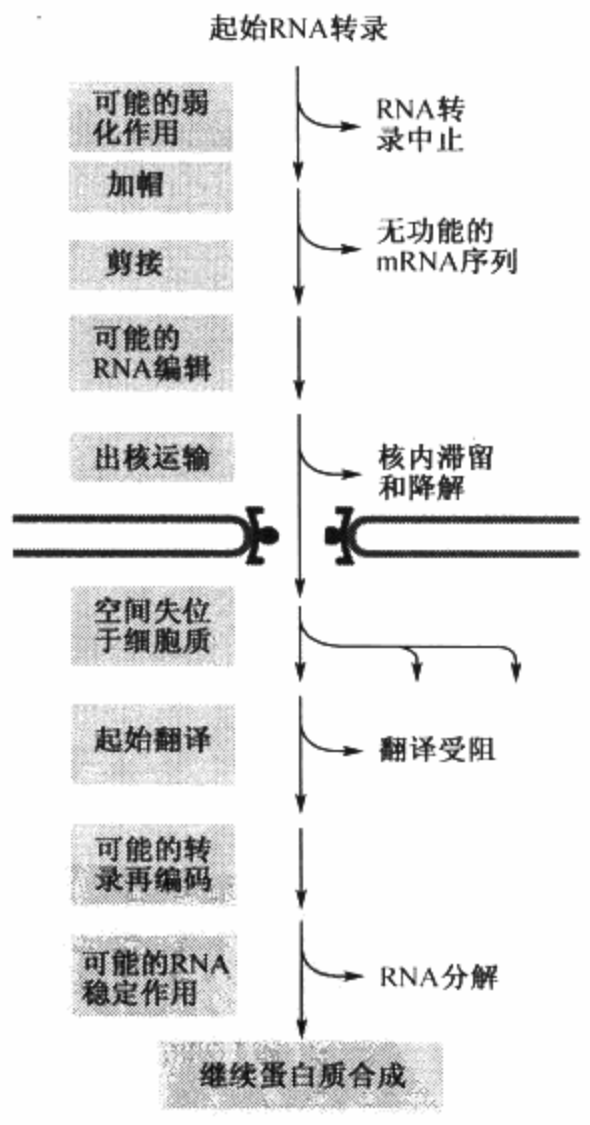


图 7-87　基因表达可能的转录后调控
对于任何一个基因而言，这些调控中可能只有几个很重要

转录弱化作用引起一些 RNA 分子提前终止

在细菌中，某些基因的表达受到转录提前终止的抑制，这个现象称为**转录弱化**。在某些这样的情况下，新生的 RNA 链会采用某种结构，使得这条 RNA 链与 RNA 聚合酶以能够中止其转录的方式相互作用。当需要该基因产物时，调节蛋白结合到新生的 RNA 链上并干扰弱化作用，从而能够转录出一条完整的 RNA 分子。

在真核生物中转录弱化可以通过不同的机制进行。一个研究的比较清楚的例子存在于 HIV（人类 AIDS 病毒）。一旦该病毒整合到人类基因组中，就由人类细胞的 RNA 聚合酶Ⅱ转录病毒 DNA（图 5-73）。然而，宿主聚合酶常常在整合完几百核苷酸转录物之后终止转录（由于尚不清楚的原因），因而不能有效地转录出整个病毒基因组。当病毒处于最佳生长条件以后，就由病毒编码的称作 Tat 的蛋白质阻止提前终止，该蛋白质结合新生的含有一个凸出碱基 RNA 中一种特殊的茎-

环结构。一旦结合上这个特殊 RNA 结构以后（称作 Tar），Tat 就与许多细胞蛋白质装配起来，这些蛋白质使 RNA 聚合酶能够持续转录。这些细胞蛋白质至少有一部分的正常作用是，当 RNA 聚合酶转录正常的细胞基因时，它会防止暂停和提前终止。真核基因常常含有长内含子；为了有效地转录基因，RNA 聚合酶Ⅱ不能在碰巧促进暂停的核苷酸序列位置处停留。因此 HIV 明显采用了一种常见的细胞机制，使其基因组转录能由单个病毒蛋白质进行调控。

选择性 RNA 剪接能从同一个基因产生不同形式的蛋白质

第 6 章讨论到，许多真核基因的转录物通过 RNA 剪接截短，在此过程中内含子序列从 mRNA 前体中除去。我们看到一个细胞能以不同的方式剪接“前”转录物，因而能从同一基因产生不同的多肽链——这个过程称作**选择性 RNA 剪接**（图 6-27 和图 7-88）。大部分高等真核基因（据估计，至少有 1/3 的人类基因）以这种方式产生多种蛋白质。

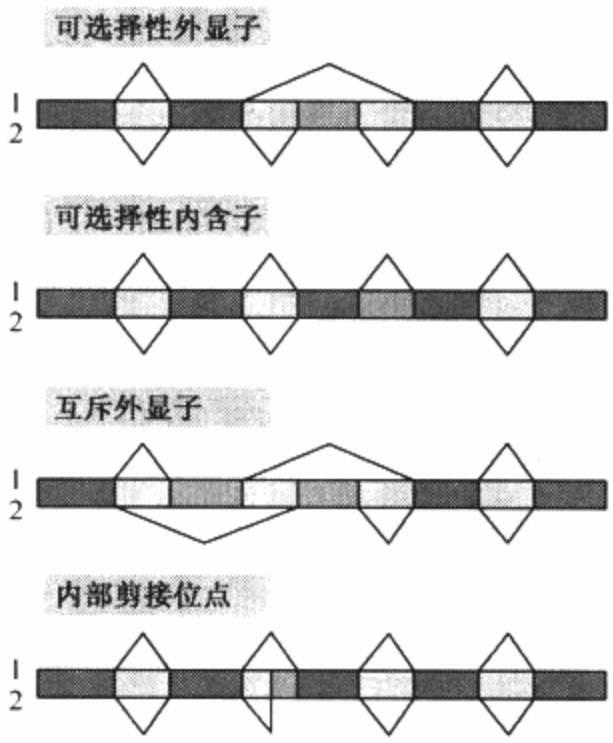


图 7-88　4 种选择性 RNA 剪接模式

在每种情况下，一种 RNA 转录物以两种选择性的方式进行剪接，生成两种不同的 mRNA（1 和 2）。深蓝色盒子标记两个 mRNA 中都保留下来的外显子序列。浅蓝色标记只在其中一个 mRNA 中存在的可能的外显子序列。这些盒子以红线相连来表示除去内含子序列（黄色）的地方（摘自 A. Andreadis, M. E. Gallego, and B. Nadal Ginard, Annu. Rev. *Cell Biol*. 3:207～242,1987）

当转录物中多个位置上存在不同的剪接可能性时，单个基因就能产生许多不同的蛋白质。在极端的情况下，一个果蝇基因能通过选择性剪接从单个基因合成多达 38 000 个不同的蛋白质（图 7-89），虽然到目前为止这些蛋白质中只有一小部分能通过实验方法观察到。考虑到果蝇基因组有近 14 000 个已鉴定的基因，很明显一种生物的蛋白质复杂度可以大大超过它的基因数量。这个例子也说明将基因数目与生物复杂度等同的危险性。例如，选择性剪接在单细胞芽殖酵母中较少发生，但在果蝇中却很常见。芽殖酵母大约有 6200 个基因，其中只有 327 个能进行剪接并且几乎所有这些剪接基因只有一个内含子。如果说果蝇的基因数只有酵母的 2～3 倍，就会大大低估了这两个基因组之间复杂度的差别。

在有些情况下出现选择性剪接是由于存在内含子序列多义性：标准的去除内含子序列的剪接体机制（讨论见第 6 章）不能区分两个或多个 5′和 3′剪接位点产生的全部选择性配对，所以在对不同的转录物上随机产生不同的选择。哪里发生这种组成性选择性剪接，就在表达该基因的所有细胞中产生该基因编码的多个蛋白质。

然而，在许多情况下选择性 RNA 剪接是可调型的调而不是组成性的。在最简单的例子中，可调型剪接用于将无功能蛋白质的合成转变为有功能蛋白质的合成。例如，催化果蝇 P 元件转座的转座酶，在生殖细胞中以有功能的形式合成，在果蝇体细胞中则以无功能的形式合成，从而使 P 元件能在整个基因组扩散，而不会损伤体细胞（图 5-70）。经研究发现转座子活性的差别是因为只在生殖细胞中转移的转座酶 RNA 中存在一个内含子序列。

除将有功能蛋白质的合成转变成无功能蛋白质合成之外，RNA 剪接的调控能在不同细胞类型中，根据细胞的需要产生不同版本的蛋白质。例如，原肌球蛋白在不同类型的细胞中以特定

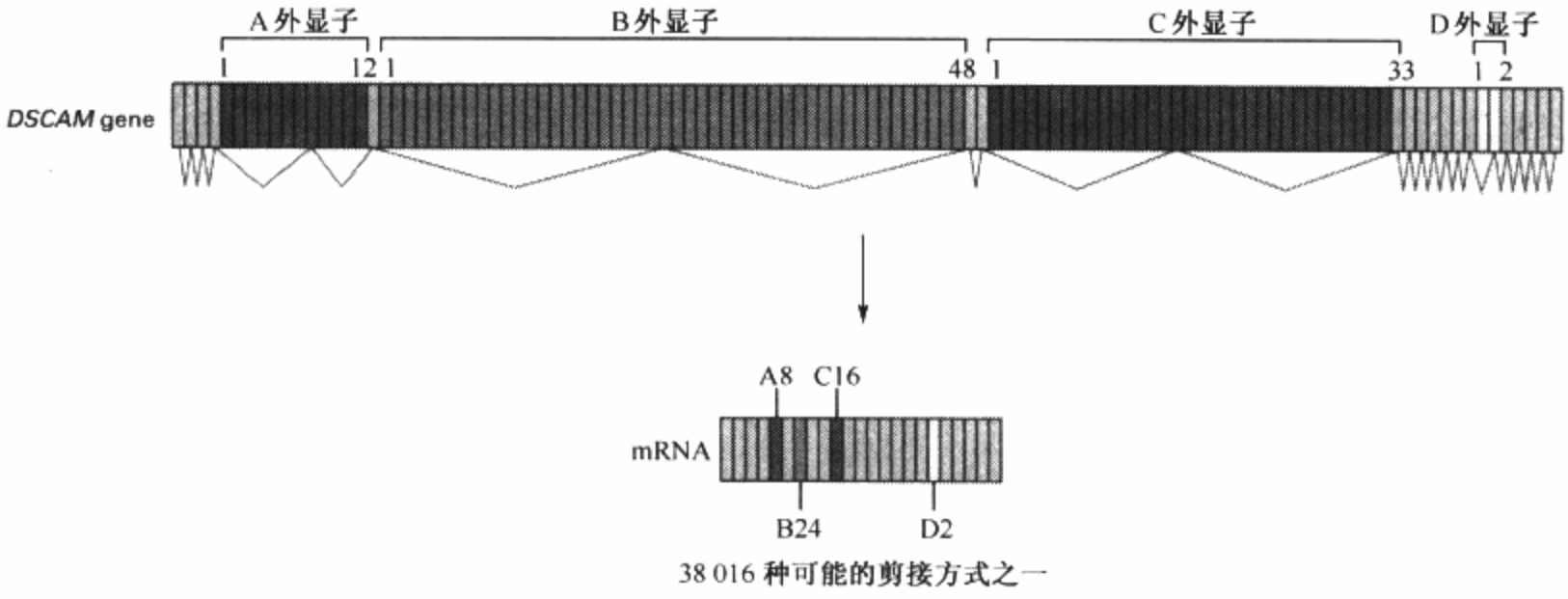


图 7-89　果蝇 *DSCAM* 基因 RNA 转录物的选择性剪接

DSCAM 蛋白是在发育的神经系统中帮助将生长锥引向其正确发育目标的轴突导向受体。最终的 mRNA 含有 24 个外显子，其中 4 个（用 A、B、C、D 表示）作为一组选择性外显子存在于 *DSCAM* 基因中。在每个 RNA 中，含有 12 个可选择外显子 A 中的一个（红色），48 个可选择外显子 B 中的一个（绿色），33 个可选择外显子 C 中的一个（蓝色），2 个可选择外显子 D 中的一个（黄色）。如果所有可能的剪接组合都得到利用，从理论上讲 *DSCAM* 基因可以产生 38 016 种不同的蛋白质。图中只显示了众多的剪接模式中的一种（用红线以及其下面的成熟 mRNA 表示）。每种变体 DSCAM 蛋白折叠成大致相同的结构（大多数都是一系列的免疫球蛋白类似胞外结构域，与跨膜区域相连）（图 24-71）。但是根据不同的剪接模式，该结构域的氨基酸序列将不同。可能该受体多样性有助于形成复杂的神经回路，但许多 DSCAM 确切的特性与功能还不清楚（摘自 D. L. Black，*Cell* 103：367～370，2000）

的形式合成（图 6-27）。许多其他蛋白质的细胞类型特异的形式也通过相同的方式合成出来。

RNA 剪接既可以受到负调控，由一个防止剪接装置接触 RNA 上特定剪接位点的调节分子进行调控；也可以受到正调控，由一个帮助引导剪接装置到达其他情况下被忽略的某个剪接位点的调节分子进行（图 7-90）。对于果蝇转座酶，重要的剪接事件在体细胞中被负调控阻断。

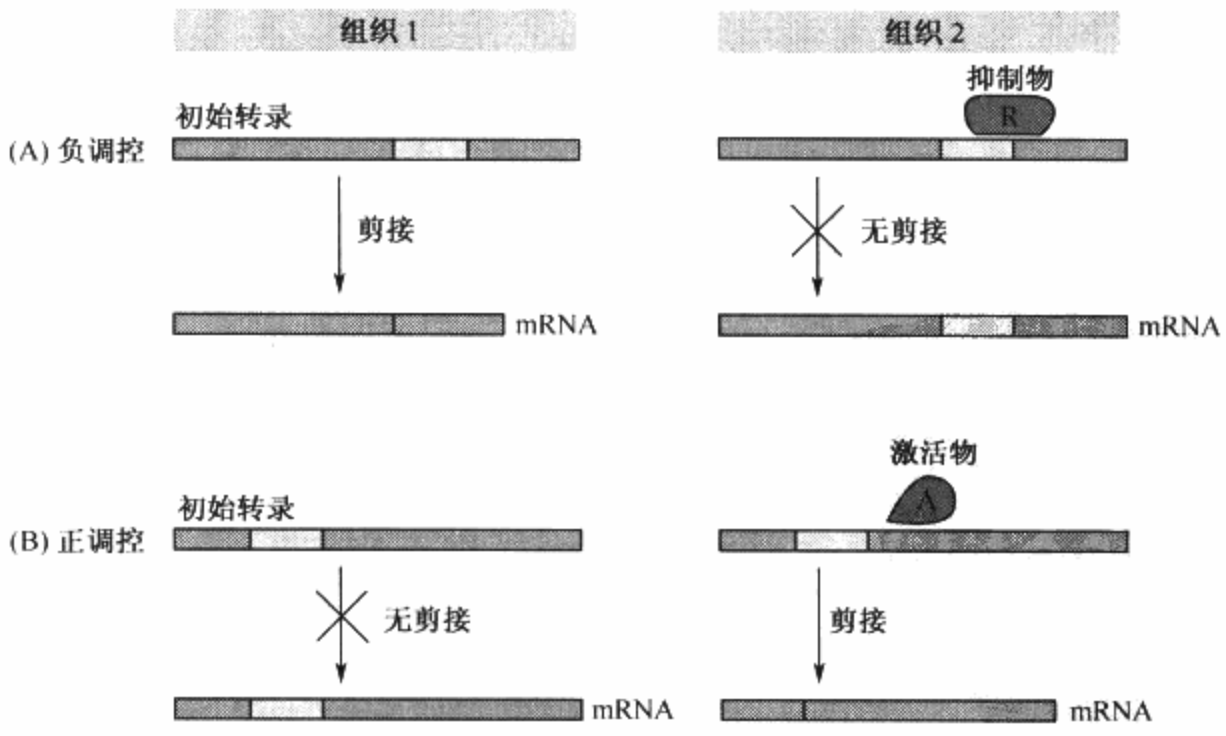


图 7-90　选择性 RNA 剪接的正调控和负调控

（A）对于负调控，阻遏蛋白结合组织 2 中的初始 RNA 转录物，因而阻止剪接装置除去内含子。（B）对于正调控，没有激活蛋白的帮助，剪接装置不能有效地除去特定的内含子序列

由于 RNA 剪接的柔性，一个强剪接位点的阻断常常会暴露一个“弱”位点，并导致产生不同剪接模式。同样，激活一个次佳的剪接位点能够通过抑制一个竞争剪接位点引发选择性剪接。因此可以认为前体 mRNA 分子的剪接是一个位于竞争剪接位点之间精细的天平——一种容易被调节蛋白倾斜的天平。

自从发现选择性剪接以后，已不得不对基因的定义做出修定

有关真核基因含有内含子且它们的编码序列能以多种方式组织起来的发现，引出了与基因定义有关的新问题。在 20 世纪 40 年代早期，根据在链胞霉属真菌中进行的生化遗传学研究工作，最早用分子生物学语言定义了基因。直到那时，还将基因定义为基因组的一个区域，它们在减数分裂时作为一个单位分离并产生确定的表型特征。例如，在果蝇中白色或红色的眼，或者豌豆中光滑或皱皮的种子。有关链胞霉属的研究表明大多数基因对应于基因组中指导合成某种酶的一个区域，这导致产生一个基因编码一条多肽链的假说，该假说被随后的研究证明是富有成效的；随着在 20 世纪 60 年代对基因表达的机制了解的越来越多，开始将基因视为一段转录成编码了一条多肽链的 RNA 的 DNA（或者一条结构 RNA 如 tRNA 或 rRNA）。假设从任何一条 DNA 序列转录而来的 RNA 都指令了一条多肽链的话，那么 20 世纪 70 年代断裂基因和内含子的发现可以容易地被最初的基因定义所容纳。但现在清楚了高等真核细胞中许多 DNA 序列通过选择性 RNA 剪接能合成一组不同但是相关的蛋白质。那么基因该如何定义呢?

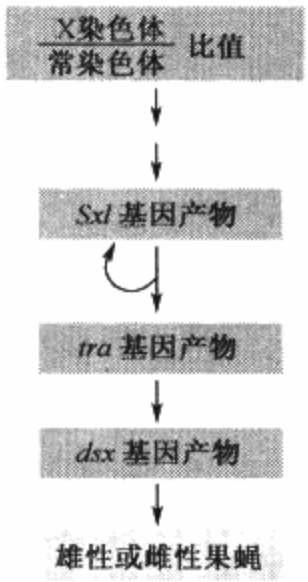


图 7-91　果蝇的性别决定
根据 X 染色体/染色体比值，图中显示的基因产物以连续的级联反应发挥作用来决定果蝇的性别。根据这些基因突变失活时产生的表型，分别将它们称为 *sex-lethal*（*Sxl*）、*transformer*（*tra*）、*doublehesex*（dsx）。这些基因产物的功能是将有关 X 染色体/常染色体比值的信息传递到其他产生性别相关表型的基因。这些其他基因作为两套可选择的形式发挥作用：那些指令雌性特征的基因以及那些指令雄性特征的基因（图 7-92）

在那些比较罕见的情况下，两种不同的真核蛋白质从一个转录单位合成出来，那么这两种蛋白质被认为由染色体上重叠的不同基因合成。然而，将选择性剪接产生的大部分蛋白质变体形式看成是来自于重叠的基因似乎不必要地复杂化了。一个更合理的定义是将最早的定义加以修改，认为基因是任何作为一个单位进行转录并编码一组密切相关的多肽链（蛋白质异型体）的 DNA 序列。这个基因定义也容纳了那些编码蛋白质变体的 DNA 序列，它们由 RNA 剪接以外的转录后加工过程产生，如翻译读码框移位（图 6-78）、可调型 poly-A 添加以及 RNA 编辑。

在果蝇中性别决定依赖于一系列可调型 RNA 剪接

现在我们来看看可调型 RNA 剪接的一个了解得最清楚的例子。在果蝇中决定果蝇是发育成雄性还是雌性的主要信号是 X 染色体/常染色体的比值。X 染色体/常染色体比值为 1 的个体（通常 2 条 X 染色体和 2 套常染色体）发育成雌果蝇，而该比值为 0.5 的个体（通常 1 条 X 染色体和 2 套常染色体）发育成雄果蝇。在发育早期对此比值进行评估，此后由每个细胞记住。3 个重要的基因产物将此比值的信号传递到许多其他指令雄性和雌性特征的基因（图 7-91）。如图 7-92 所解释的，果蝇性别决定依赖于一组可调型 RNA 剪接事件，它们参与合成这 3 种基因产物。

虽然果蝇性别决定提供了基于 RNA 剪接的级联调节系统的一个了解最清楚的例子，但还不清楚果蝇为什么采用这种策略。其他

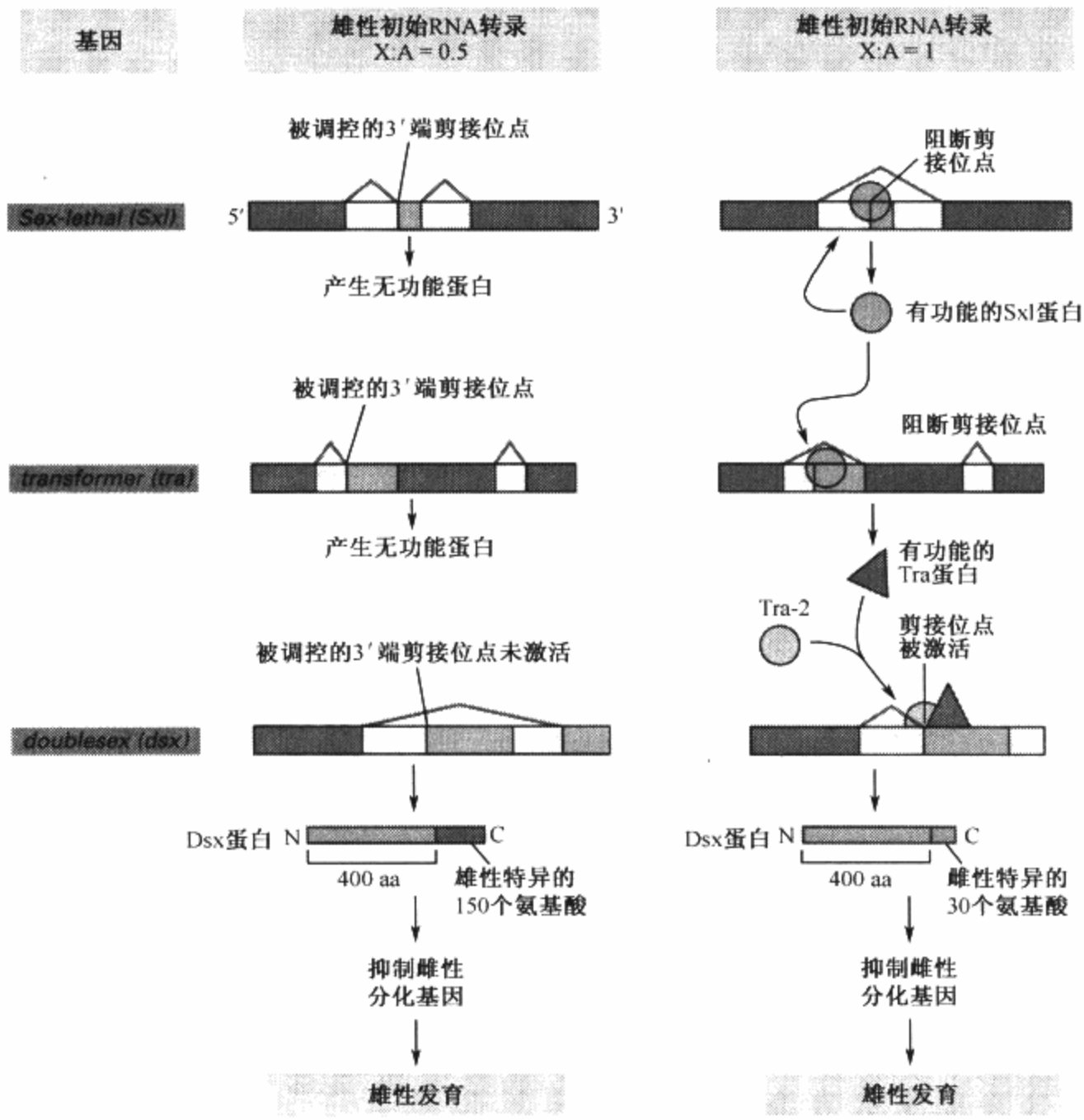


图 7-92　通过选择性 RNA 剪接决定果蝇性别的基因表达过程中的级联变化

X 染色体/常染色体比值为 0.5 时引起雄性果蝇发育。发育成雄性是默认的 *Sxl* 和 *tra* 都发生转录的途径，但是 RNA 组成性剪接只产生无功能的 RNA 分子，并且 *dsx* 转录物剪接后生成的蛋白质将指令雌性特征的基因关闭。X 染色体/常染色体比值为 1 时，通过短暂活化 *Sxl* 基因中的启动子，引发胚胎中雌性分化途径，而 *Sxl* 基因引起合成一类特殊的 *Sxl* 转录物，再经过组成性剪接后生成有功能的 Sxl 蛋白。Sxl 是具有两个作用位点的剪接调节蛋白：①它结合组成性合成的 *Sxl* RNA 转录物，引起雌性特异的剪接，再由这种剪接继续合成有功能的 Sxl 蛋白。②它结合组成性合成的 *tra* RNA 并引起该转录物的选择性剪接，现在该剪接生成活性 Tra 调节蛋白。Tra 蛋白与组成性合成的 Tra-2 蛋白共同作用，产生 *dsx* 转录物雌性特异的剪接形式；剪接产物编码了 Dsx 蛋白的雌性形式，由它将指令雄性特征的基因关闭。这条途径中的成员最初都是通过研究性别发育发生改变的果蝇突变体得以鉴定出来。例如，*dsx* 基因（*doublehesex*）名字的由来是因为观察到缺少该基因产物的果蝇既表达雄性特异的特征，又表达雌性特异的特征。注意虽然 Sxl 和 Tra 蛋白都结合特异的 RNA 位点，但 Sxl 是一个阻遏蛋白，它发挥副作用，阻断剪接位点，而 Tra 蛋白是一个激活蛋白，它发挥正作用，诱导剪接（图 7-90）。Sxl 结合富含嘌呤核苷酸的片段，它是标准剪接共有序列的一部分（图 6-28）并阻止正常剪接因子，U2AF（图 6-29）的进入。Tra 结合外显子中的特异 RNA 序列并激活最适度以下的剪接信号

生物（如线虫）采用完全不同的方案做出性别决定——一种基于转录和翻译调控的方案。此外，果蝇雄性决定途径要求不断合成许多无功能的 RNA 分子，这看起来是不必要的浪费。一种推测认为这种 RNA 剪接级联系统利用了一种古老的调控装置，它从进化早期阶段遗留

下来，在那时 RNA 是占优势地位的生物学分子，并且基因表达的调控几乎不得不完全基于 RNA-RNA 相互作用（讨论见第 6 章）。

RNA 转录物切割和 poly A 加成的位点发生改变能使一个蛋白质的 C 端发生改变

第 6 章我们看到，真核 mRNA 分子的 3′端不是由 RNA 聚合酶催化的 RNA 合成终止形成的，而是在转录还在延伸时，由其他的因子催化 RNA 切割反应生成的（图 6-37）。细胞能调控这种切割的位点，以改变所生成蛋白的 C 端。

一个经过仔细研究的例子是在 B 淋巴细胞发育期间发生的抗体分子从膜结合型向分泌型转变。在 B 淋巴细胞生命历史早期，它产生的抗体锚定在细胞质膜上，在这里作为抗原的受体。抗原刺激引起 B 淋巴细胞增殖并开始分泌它们的抗体。分泌的抗体形式与膜结合的形式除 C 端不同外其他方面都相同。在蛋白质的 C 端这一部分，膜结合型有一长串疏水氨基酸，横跨膜脂双层，而分泌型则是一串短得多的疏水氨基酸。从膜结合型向分泌型转换因而需要 mRNA 3′端存在不同的核苷酸序列；初始 RNA 转录物长度的改变产生这种差别，如图 7-93 所示，RNA 转录物长度的改变又因 RNA 剪接位点的改变引起，后者因 CstF 的一个亚基浓

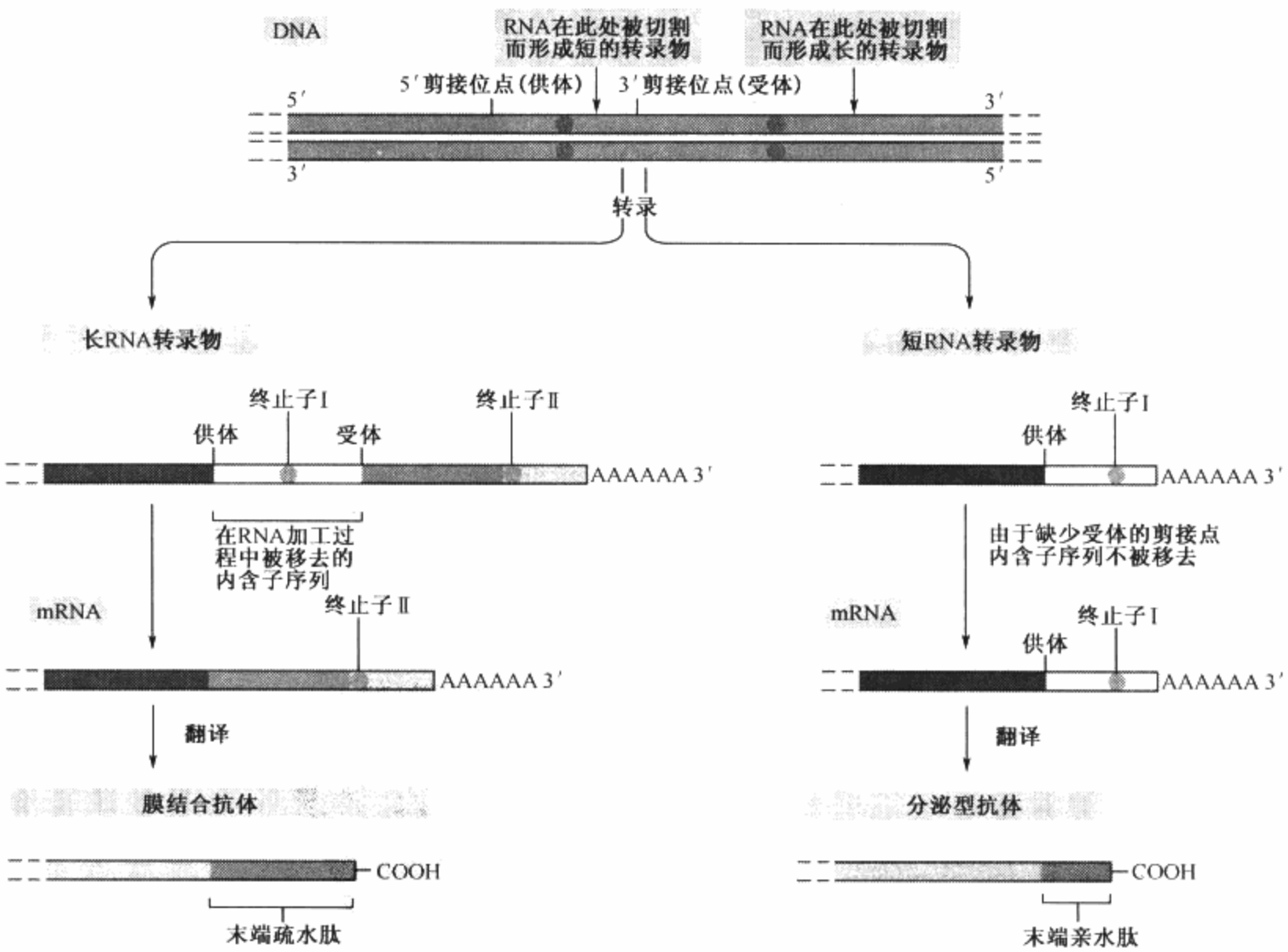


图 7-93　对 RNA 切割和 poly A 加成位点的调控决定了抗体是分泌出去还是保持与膜结合

在未受刺激的 B 淋巴细胞中（左边），合成长 RNA 转录物，其 3′端的内含子序列通过 RNA 剪切除去以生成编码膜结合抗体分子的 mRNA 分子。相比之下，抗原刺激（右边）以后，初始 RNA 转录物在位于最后一个外显子序列前面的剪接位点上游切割。结果，有些从长转录物去除的内含子序列保留下来作为短转录物中的编码序列。这些是编码了分泌型抗体分子亲水 C 端部分的核苷酸序列

度增高而引起，CstF 是结合 RNA 切割和 poly A 加成位点富含 G/U 序列并促进 RNA 切割的蛋白质（图 6-37 和图 6-38）。转录抗体基因的 RNA 聚合酶遇到的第一个切割-poly A 加成的位点不是最合适的位点，在未受刺激的 B 淋巴细胞中通常被跳过，结果生成较长的 RNA 转录物。当抗体刺激引起 CstF 浓度增加时，切割就在此非最适位点进行，生成较短的转录物。通过这种方式，一个通用 RNA 加工因子浓度的变化能对较少数量的基因产生特定的影响。

RNA 编辑能改变 RNA 信息的含义

细胞运用的分子机制不断让人感到奇怪。一个例子就是**RNA 编辑**过程，它改变了mRNA转录出来以后的核苷酸序列。在第 6 章，我们已经看到 rRNA 和 tRNA 在转录后发生修饰。这一部分我们会看到有些 mRNA 以改变它们携带的编码信息的方式进行修饰。最显著的 RNA 编辑形式在编码锥虫线粒体蛋白的 mRNA 转录物中发现。其中有一个或多个 U 核苷酸插入（或去除，较少发生）到转录物经过选择的区域，引起最初的可读框和序列发生较大的改变，因而改变了信息的含义。对于有些基因而言，编辑如此广泛地进行，以至于在成熟 mRNA 中超过一半的核苷酸是在编辑过程中插入的 U 核苷酸。指令最初的 RNA 转录物究竟将如何改变的信息包含在一组单独转录的长 40～80 个核苷酸的 RNA 分子中。这些所谓的指导 RNA 5′端序列与将要发生编辑的转录物的一个末端区域互补；这个 5′端后面接着是一段序列，指令将要插入转录物的一组核苷酸，继而这段序列后面是重复的 U 核苷酸。编辑机制非常复杂，在此过程中指导 RNA 3′端的 U 核苷酸直接转移到转录物中，如图 7-94 所示。

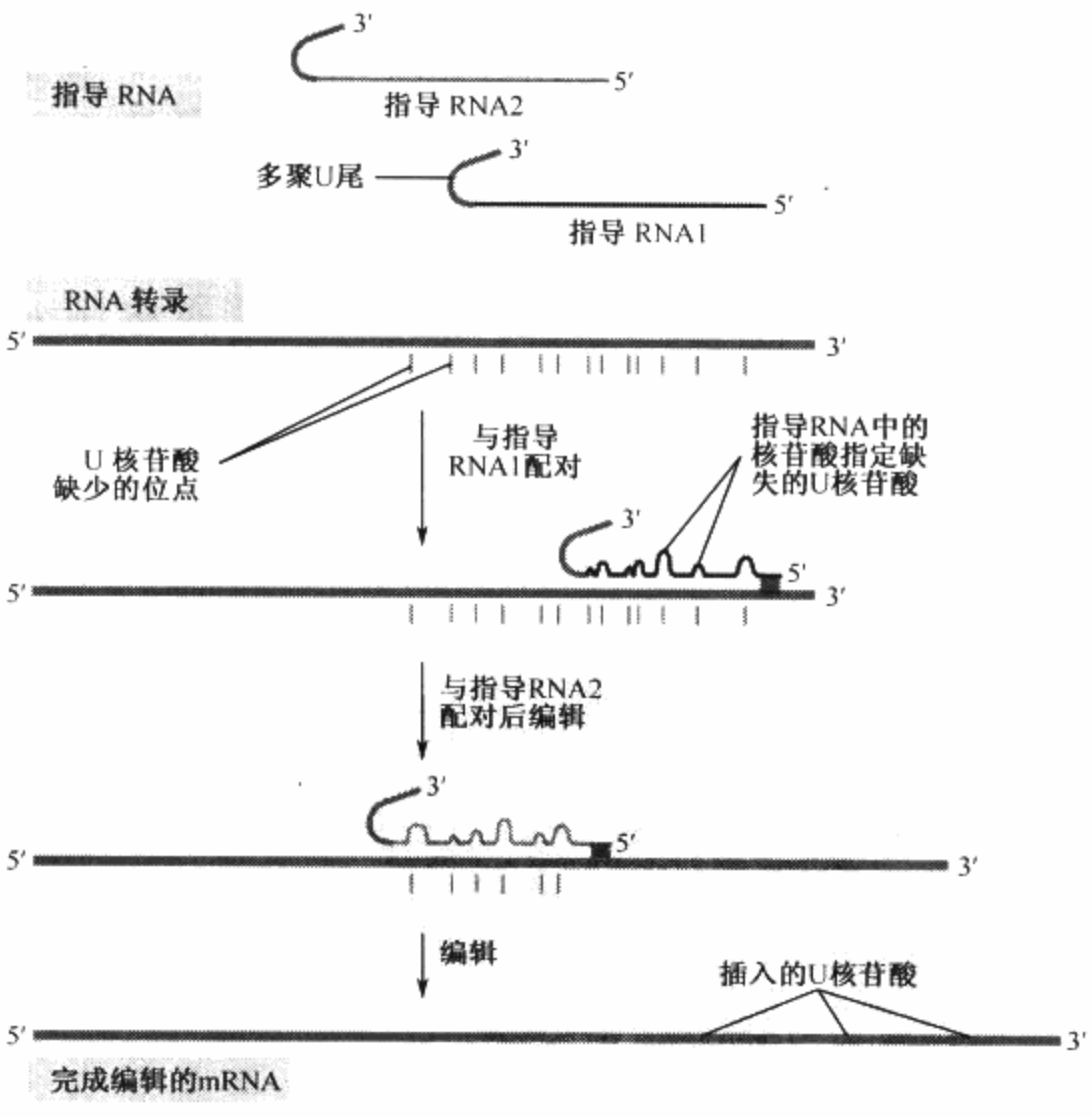


图 7-94　发生在锥虫线粒体中的 RNA 编辑

指导 RNA 的 3′端含有一段多聚 U，它为 RNA 转录物上与指导 RNA 错配的位点提供 U 核苷酸。因此多聚 U 尾随着编辑的进行会变短（未显示）。编辑一般从 3′端附近开始，向 RNA 转录物 5′端前进，如图所示，因为大多数指导 RNA 的末端锚定序列只能与被编辑的序列配对

许多植物线粒体中也发现了广泛的 mRNA 序列编辑，其中几乎每个 mRNA 都以某种程度进行了编辑。然而在这种情况下，RNA 碱基从 C 变为 U，不发生核苷酸插入或删除。通常编辑影响一个 mRNA 中许多的 C，从而改变了这个 mRNA 编码的 10%或更多的氨基酸。

至于锥虫和植物线粒体为什么要利用这些广泛的 RNA 编辑，我们只能进行推测。似乎最合理的观点基于一个先决条件，就是线粒体拥有一个原始的遗传系统，其中不存在其他调控系统的作用可能性。存在证据表明编辑受到调节，以便在不同条件下生成不同的 mRNA，这样 RNA 编辑就可以看成改变基因表达的原始方式，也许是在很古老的细胞中运作的机制的遗迹，在这些细胞中大部分催化作用由 RNA 分子而不是蛋白质执行。

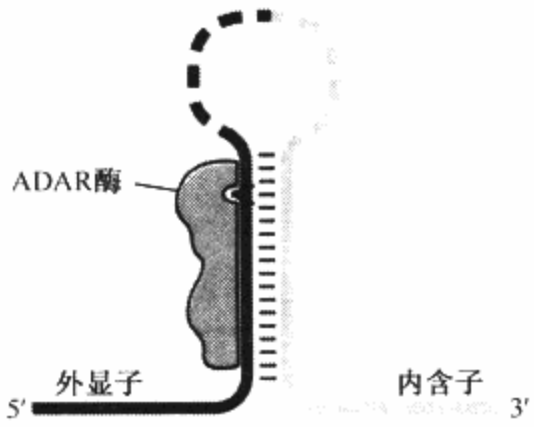


图 7-95　哺乳动物中 A→I 的 RNA 编辑机制
编辑发生的位置由同一 RNA 分子中携带的 RNA 序列提供信号。通常与编辑位置互补的序列存在于内含子中，结果生成的双链 RNA 吸引 A→I 编辑酶 ADAR。小鼠和人类有 3 种 ADAR 酶：ADAR1 在肝脏中需要，用于正确的血红细胞的发育；ADAR2为脑的发育所需（如文中所述）；ADAR3 的作用还不清楚

哺乳动物中存在一种更加受限制的 RNA 编辑形式。最重要的类型之一就是腺嘌呤酶促脱氨基后生成次黄嘌呤（图 6-55），这个过程在有些 mRNA 中经过选择的位点发生。在有些情况下，这种修饰改变了 RNA 的剪接模式；在另外一些情况下，它改变了密码子的含义。由于次黄嘌呤与胞嘧啶配对，所以由 A 变为 I 的编辑生成一个氨基酸序列发生改变的蛋白质，这种编辑由称作 ADAR（作用于 RNA 的腺嘌呤脱氨酶）的蛋白酶催化进行；这些酶识别将要发生编辑的位点之间形成的双链 RNA 结构和位于同一 mRNA 分子其他位置，通常是在 3′内含子中的一段互补序列（图 7-95）。大脑中一种编码递质门控离子通道的前体 mRNA 中，存在一个特别重要的 A→I 编辑例子。发生一次编辑将谷氨酸变成精氨酸；受影响的氨基酸位于离子通道内壁，编辑改变了通道的 Ca^{2+} 通透性。通过删除相关的 ADAR 基因，证明了这种编辑在小鼠中很重要。小鼠突变体容易癫痫发作并在断奶期间或之后死亡。如果将编码门控离子通道的基因突变，直接生成编辑后的蛋白质形式，缺少 ADAR的小鼠就能正常发育，表明离子通道 RNA 的编辑对于正确的大脑发育通常很关键。小鼠和人类有许多其他的 ADAR 基因，在小鼠中删除它们之中的一个会导致出生前小鼠胚胎的死亡。由于这些小鼠在合成红细胞前体时存在许多缺陷，所以 RNA 编辑可能对于造血系统的发育也很关键。

C→U 编辑也可以在哺乳动物中观察到。其中一个例子，即在载脂蛋白 B mRNA 的编辑中，C 转变为 U 产生一个终止密码子，从而导致这个以组织特异的方式合成的巨大蛋白质缩短了。为什么哺乳动物中一定要存在编辑还是一个谜。一种观点认为它在进化中出现，以便更正基因组中的“错误”。另一种观点认为它又提供了一种从同一基因合成多种相关蛋白质的方式。第三种观点认为编辑只是由随机突变产生的大量有害的临时代用装置之一，由于它们偶然提供了有用的作用所以得以长期存在。

从细胞核转运 RNA 可以受到调控

曾有人估计哺乳动物合成的全部 RNA 中约有 1/20 曾离开过细胞核。第 6 章我们看到大部分哺乳动物 RNA 分子经历广泛的加工，残余的 RNA 片段（切除的内含子和切割/poly A 位点 3′端的 RNA 序列）在细胞核被降解掉。不完全加工和其他情况下损伤的 RNA 也最终在

细胞核中作为 RNA 合成质量控制体系的一部分被降解掉。这种降解由**外核体**执行，它是一个巨大的蛋白质复合体，含有多种不同的 RNA 外切核酸酶作为亚基。

第 6 章描述过，RNA 分子从细胞核的输出过程一直延迟，直到加工完成以后才进行。因此从理论上讲，任何阻止特定 RNA 分子上完成 RNA 剪接的机制都能阻断 RNA 从细胞核出来。这个特征是一个了解最清楚的 mRNA **可调型核转运**例子的理论基础。这种转运在引起艾滋病的 HIV 中进行。

HIV 是一种反转录病毒（一种 RNA 病毒），一旦进入细胞，它就指导合成其基因组的双链 DNA 拷贝，然后这个拷贝插入到宿主基因组中（图 5-73）。一旦插入以后，病毒 DNA 由宿主细胞 RNA 聚合酶Ⅱ以一个长 RNA 分子转录出来。然后这个转录物以许多不同的方式进行剪接，生成超过 30 种不同种类的 mRNA，继而这些 mRNA 再翻译成各种不同的蛋白质（图 7-96）。为了产生后代病毒，必须将完整的未剪接的转录物从细胞核转运到细胞质中，并在这里将它包装到病毒衣壳中（图 5-73）；此外，许多 HIV mRNA 以一种方式进行选择性剪接，使得它们仍然携带完整的内含子。因此宿主细胞对未剪接 RNA 输出细胞核（以及随后进行的降解）的阻断，给 HIV 造成一个特殊的难题，但它以一种巧妙的方式加以克服。

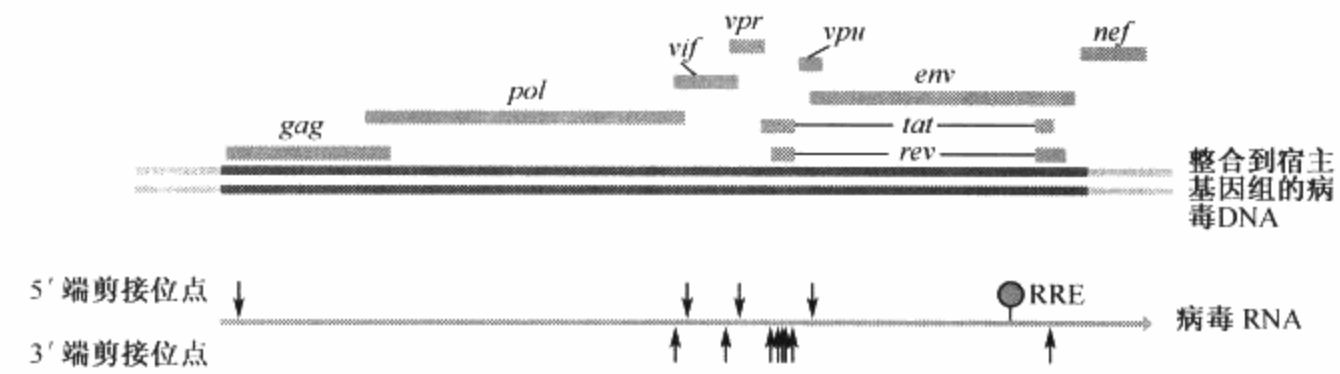


图 7-96　HIV（人类艾滋病病毒）紧密的基因组

绿色显示 9 个 HIV 基因的位置，红色双线表示整合到宿主 DNA（灰色）中的病毒基因组 DNA 拷贝。注意许多基因的编码区域重叠，且 *rev* 和 *tat* 的编码区被内含子打断。图底部的蓝线代表病毒 DNA 的前体 mRNA 转录物，其中显示了所有可能的剪接位点的位置（箭头）。病毒转录物存在许多可选择的剪接方式，如 *env* mRNA 保留了 *tat* mRNA 和 *rev* mRNA 中剪接掉的内含子。蓝色球棒代表 Rev 响应元件（RRE）。它是一段长 234 个核苷酸的 RNA，折叠成精确的结构，Rev 识别这个较大结构中特定的发夹（图 6-94）。*gag* 基因编码的蛋白质切割成几个更小的蛋白质并由它们形成病毒衣壳。*pol* 基因编码的蛋白质经过切割产生反转录酶（它将 RNA 转录成 DNA），以及参与整合病毒基因组（以双链 DNA 形式）到宿主基因组中的整合酶。*pol* 由起始于 *gag* 的核糖体移码翻译合成（图 6-78）。*env* 基因编码包被蛋白（图 5-73）。Tat、Rev、Vif、Vpr、Vpu 和 Nef 都是具有多种功能的小蛋白质。例如，Rev 调节核输出（图 7-97），Tat 调节转录跨整合的病毒基因组进行延伸

HIV 编码一种蛋白质（称作 Rev），能结合位于病毒一个内含子中的特定 RNA 序列（称作 Rev 响应元件，RRE）。Rev 蛋白与一个核内输出受体（输出蛋白 1）相互作用，这个受体指导病毒 RNA 移动通过核孔到达细胞质，而不论是否存在内含子序列。我们将在第 12 章详细讨论输出受体的作用方式。

Rev 对核输出的调节对 HIV 生长和病原发生产生许多重要的影响。除了确保特定的未剪接 RNA 的核输出以外，它还将病毒感染分成早期（此时 Rev 从完成剪接的 RNA 翻译得到，而含有内含子的 RNA 保留在细胞核并被降解）和晚期（此时由于 Rev 的作用将未剪接的 RNA 输出）。这个时间的划分通过大致按所需的先后顺序提供基因产物（图 7-97），从而有助于病毒复制。还可能 Rev 的调控帮助 HIV 病毒获得潜伏性，在此情况下 HIV 基因组整合到

宿主基因组中，但病毒蛋白质的合成暂时中止了。在它最初进入宿主细胞以后，如果条件变得对病毒转录和翻译不利，Rev 就以很低的水平合成，以至于不能促进未剪接的 RNA 输出。这种情况使病毒生长周期停止，当病毒复制的条件改善以后，Rev 的水平增加，病毒才能进入复制周期。

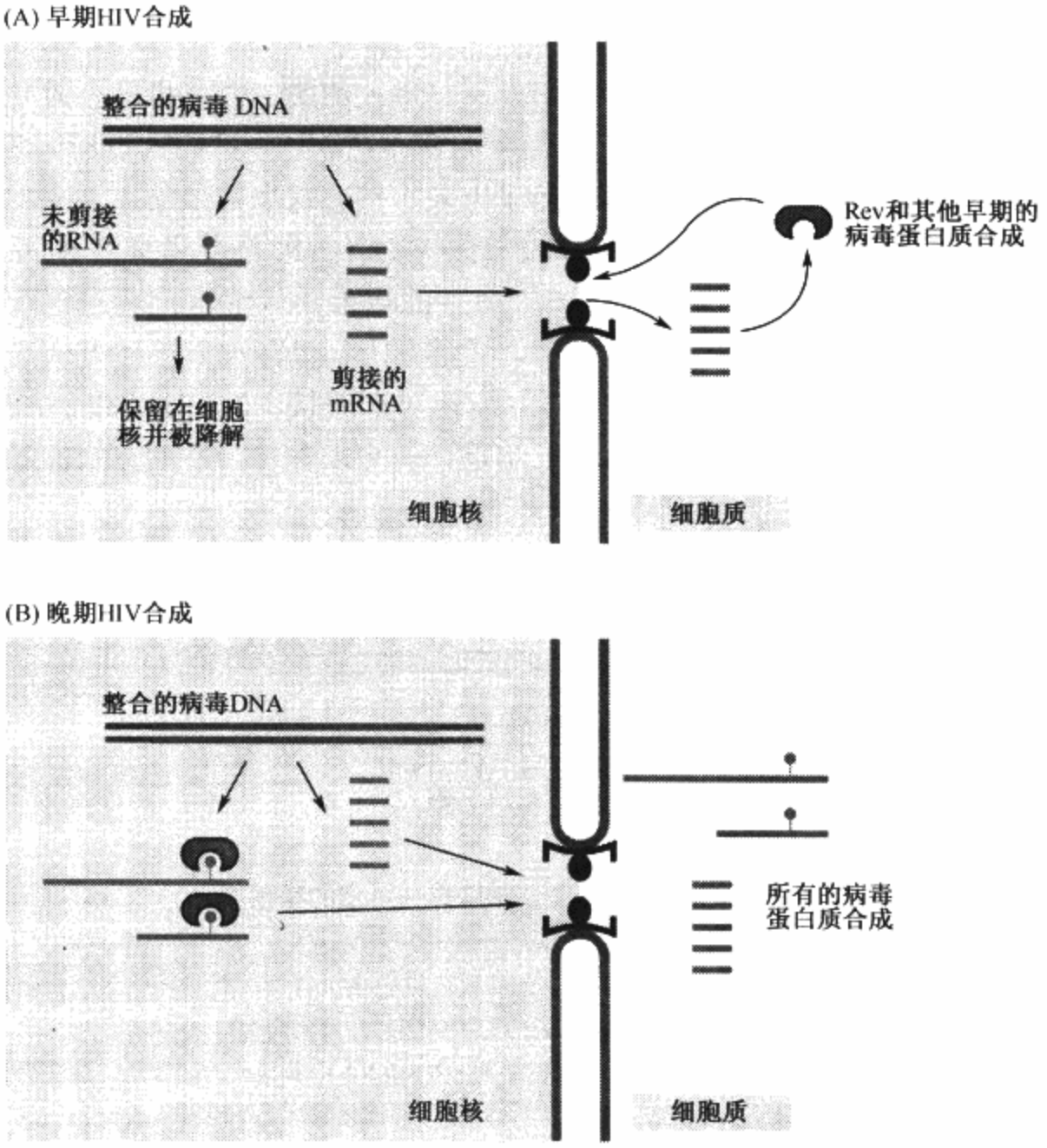


图 7-97　HIV Rev 蛋白调节核输出

在 HIV 感染早期（A），只有完全剪接的 RNA（它们含有 Rev、Tat 和 Nef 的编码序列）从核输出并进行翻译。一旦累积了足够的 Rev 蛋白并转运到细胞核（B），那么未剪接的病毒 RNA 就能从细胞核输出。许多这些 RNA 翻译成蛋白质并且全长的转录物包装到新的病毒颗粒中

有些 mRNA 定位在细胞质中的特定区域

一旦新合成的真核 mRNA 分子通过核孔进入细胞质，它通常会遇上核糖体并将它翻译成一条多肽链（图 6-40）。如果这条 mRNA 编码的蛋白质注定要分泌或在细胞表面表达，那么位于该蛋白质氨基酸末端的信号序列就会引导它到达内质网；一旦该信号序列在核糖体上出现，蛋白质分类装置的组分就会识别它，并将核糖体、mRNA 和新生蛋白质组成的整个复合体引导至内质网膜，在这里合成剩余的多肽链（讨论见第 12 章）。在其他情况下，整个蛋白质由细胞质中的游离核糖体合成，然后可能完整多肽链中的信号引导该蛋白质到达细胞的其他位点。

有些 mRNA 自身在翻译开始前被引导至特定的细胞内位置。可能对于细胞来讲，将其 mRNA 定位在该 mRNA 合成的蛋白质为细胞所需的位置是有利的。引导 mRNA 定位的信号通常位于 mRNA 分子非翻译区（UTR）的 3′端——该区域从终止蛋白质合成的终止密码子延伸到 poly A 尾开始的位置[图6-22(A)]。在果蝇卵中可观察到 mRNA 定位的一个显著例子，其中编码 *bicoid* 基因调节蛋白的 mRNA 通过附着到细胞骨架上定位于发育卵的前端。当受精引发该 mRNA 翻译时，就形成一个 Bicoid 蛋白梯度，它在指导胚胎前端部分的发育中发挥关键作用（如图 7-52 所示，并在第 21 章中有更详细的讨论）。体细胞中许多 mRNA 以类似的方式定位。例如，编码肌动蛋白的 mRNA 利用 UTR 的信号，定位到哺乳动物成纤维细胞富含肌动蛋白纤维的细胞表层。

RNA 定位已在许多生物中观察到，包括单细胞的真菌、植物和动物，它可能是细胞用来在特定位置集中蛋白质高水平合成的一种常见机制。已经发现许多不同的 mRNA 定位机制（图 7-98），但是它们全部都需要位于 mRNA 自身的特定信号，一般集中在 3′UTR(图 7-99)。

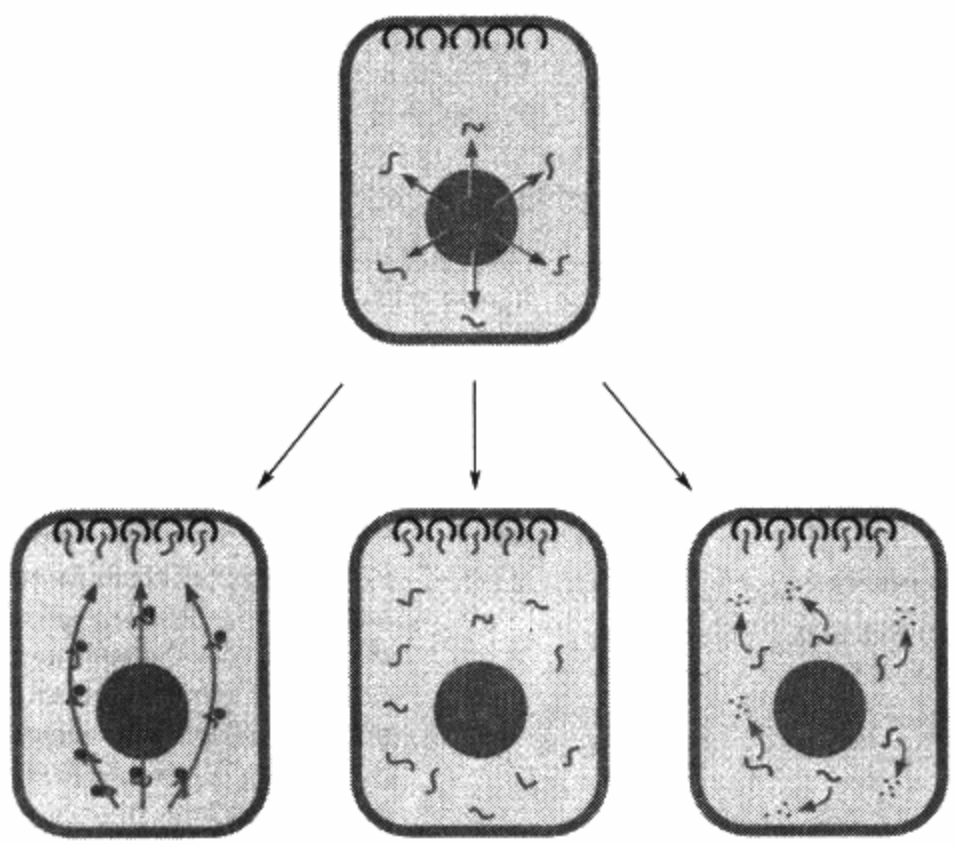


图 7-98　mRNA 定位的 3 种机制

将要定位的 mRNA 通过核孔（上面）离开细胞核。有些被定位的 mRNA（左图）结合细胞骨架传动蛋白（绿色）后向目标位置运动，如第 16 章所述。这些传动蛋白利用 ATP 水解释放的能量沿细胞骨架组分单向移动。在它们的目的地，这些 mRNA 分子由锚定蛋白（黑色）固定住。其他的 mRNA 则在细胞质随机扩散，并只是被捕获后因而在它们定位的位点上聚集（中图）。还有另一些 mRNA（右图）在细胞质中降解，除非它们通过随机扩散结合到已经定位的蛋白质复合体上，由复合体锚定 mRNA 并保护它不被降解（黑色）。每个机制都要求 mRNA 上存在特殊信号，它们通常位于 3′UTR 中（图 7-99）。在许多 mRNA 定位的情况中，其他机制阻断了 mRNA 的翻译，直到它得到正确定位为止（摘自 H. D. Lipshitz and C. A. Smibert, *Curr. Opin. Gen. Dev.* 10：476～488，2000）

第 6 章我们看到 5′帽子和 3′ poly A 尾对于有效的翻译是必需的，因此它们在同一 mRNA 分子上的存在也为翻译装置提供了信号，表明这个 mRNA 分子是完整的。刚才描述过，

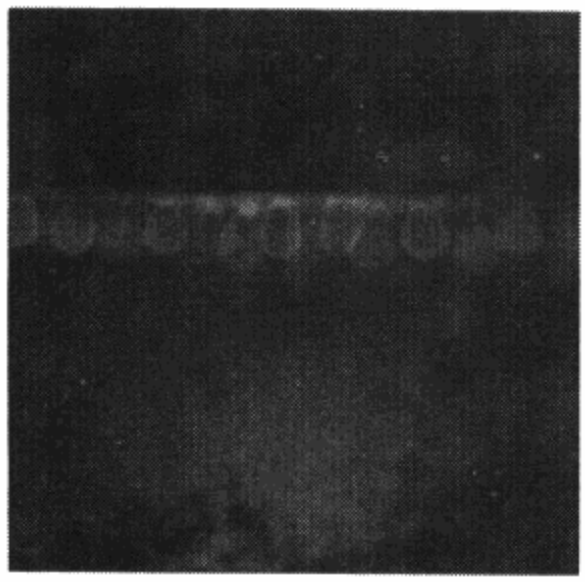


图 7-99　3′UTR 在 mRNA 定位到细胞质特殊区域过程中的重要性

该实验中，在荧光标记的 UTP 衍生物存在情况下，体外转录 DNA，从而制备两种不同的荧光标记 RNA：一种 RNA（用红色荧光染料标记）含有果蝇 hairy 蛋白的编码区以及邻近的 3′UTR；另一种 RNA（用绿色荧光染料标记）含有 hairy 蛋白的编码区，但 3′UTR 删除。在发育的一个阶段，当多个核存在于共同的细胞质中时，将这两种 RNA 混合后注入果蝇胚胎（图 7-52）。10min 后观察荧光 RNA，全长的 hairy RNA（红色）定位于核顶侧（蓝色），但缺少 3′UTR（绿色）的转录物，不能够定位，hairy 是在发育果蝇胚胎中指令定位信息的许多基因调节蛋白之一（讨论见第 21 章）。它的 mRNA 定位（该实验显示依赖于 3′UTR）据认为对于正确的果蝇发育是至关重要的（经 Simon Bullock 和 David Ish-Horowicz 许可）

3′UTR 常常含有一个“zip 密码”，它引导 mRNA 到细胞内的不同位置。在本章中，我们也将看到 mRNA 还携带一些信息，规定了每个 mRNA 在细胞质存在的平均时间，以及每个 mRNA 翻译成蛋白质的效率。从广义上讲，真核 mRNA 非翻译区类似基因的转录调控区：它们的核苷酸序列含有指令利用 RNA 的方式的信息，解释这些信息的蛋白质特异性地结合这些序列。因此，除了指令蛋白质氨基酸序列以外，mRNA 分子也富含许多其他信号。

结合 mRNA 5′和 3′非翻译区的蛋白质介导了翻译负调控

一旦合成了一个 mRNA，调控其蛋白质产物水平最常见的方式之一就是通过调控翻译起始的步骤。尽管真核生物和原核生物之间翻译起始的详细机制是不同的（我们在第 6 章看到过），但仍然利用了一些共同的基本调控策略。

在细菌 mRNA 中，常常发现一段保守的六核苷酸，Shine-Dalgarno 序列，位于起始 AUG 密码子上游几个核苷酸位置。这个序列与核糖体小亚基中的 16S RNA 形成碱基配对，从而将起始 AUG 密码正确地定位在核糖体中。由于这种相互作用对起始效率做出了主要的贡献，所以它为细菌细胞提供了一种简单的通过**翻译负调控**机制调节蛋白质合成起始的方式。这些机制一般涉及 Shine-Dalgarno 序列的阻断，要么是用结合蛋白对它进行覆盖，要么是将它整合到 mRNA 中某个碱基配对区域。许多细菌 mRNA 具有特异的翻译阻遏蛋白，它们能在 Shine-Dalgarno 序列周围结合，因而只会抑制该细菌 mRNA 的翻译。例如，有些核糖体蛋白能通过结合非翻译区的 5′端阻遏它们自身 mRNA 的翻译。这种机制只有当核糖体蛋白的合成量超过核糖体 RNA 因而不会整合到核糖体中时才发挥作用。通过这种方式，细胞能够维持用于形成核糖体的多种组分正确的数量平衡。不难猜测这个机制可能是如何进化得到的。核糖体蛋白通过结合 rRNA 中的特定位点组装成核糖体；它们中有些通过与自身 mRNA 中存

在的类似位点结合，巧妙地运用这种 RNA 结合能力来调控自身的合成。

真核 mRNA 并不含有 Shine-Dalgarno 序列。第 6 章讨论过，翻译起始位点 AUG 密码子的选择而是主要由它与 mRNA 5′端帽子靠近决定，这是核糖体小亚基结合该 mRNA 并开始扫描搜寻起始 AUG 密码子的位点。尽管翻译起始不同，真核生物也利用了翻译阻遏蛋白。它们有些结合 mRNA 5′端因而抑制翻译的起始，另外一些识别特定 mRNA 3′UTR 中的核苷酸序列并通过干扰 5′帽子和 3′ poly A 尾之间的通讯来减少翻译起始，这种通讯是有效翻译所必需的（图 6-72）。

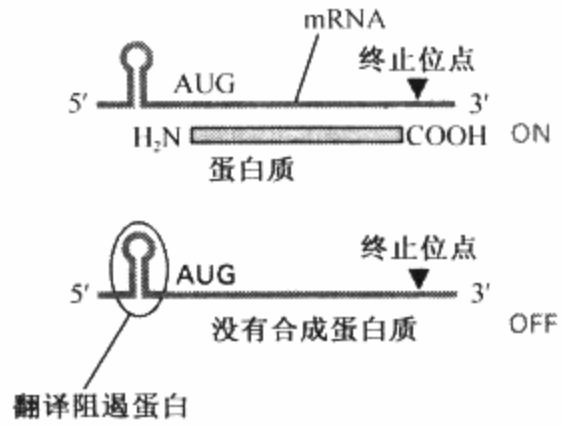


图 7-100　翻译负调控

这种调控形式由作为翻译阻遏物的序列特异性 RNA 结合蛋白介导。该蛋白质结合到 mRNA 分子上会减少 mRNA 的翻译。已经知道了这种翻译调控类型的多个实例。该幅图以细胞质中游离铁浓度增高时促使合成更多的铁蛋白（一种铁储存蛋白）的机制为模型；铁敏感翻译阻遏蛋白被称作顺乌头酸酶（图 7-105）。在其他例子中，互补的 RNA 分子而不是蛋白质，调控翻译起始。它通过形成短双螺旋 RNA 区域，阻断了 mRNA 中的关键区域

真核生物中存在一个研究得较详细的翻译负调控形式，它使得细胞内铁储存蛋白、铁蛋白的合成在细胞质中的可溶铁原子水平增高时能迅速增加。铁的调节依赖于铁蛋白 mRNA 5′前端一段约 30 个核苷酸的序列。这个铁效应元件折叠成一种茎环结构，结合一个称作顺乌头酸酶的翻译阻遏蛋白，该蛋白质阻遏下游任何 RNA 序列的翻译（图 7-100）。顺乌头酸酶是铁结合蛋白，细胞接触铁会引起它与铁蛋白 mRNA 解离，不再阻遏翻译，并且多达上百倍地增加铁蛋白的合成。

一个起始因子的磷酸化整体上调节了蛋白质合成

真核细胞减少它们的蛋白质整体合成速率以应答不同的情况，包括生长因子或营养物质的缺乏、病毒感染和温度突然升高。这种减少大多是由于翻译起始因子 eIF-2 被响应条件改变的特定蛋白激酶磷酸化而引起。

第 6 章概括了 eIF-2 的正常功能。它与 GTP 形成一个复合物，并介导甲硫氨酰起始 tRNA 结合核糖体小亚基，然后该小亚基再结合 mRNA 5′端，并开始沿 mRNA 扫描。当识别出一个 AUG 密码子时，eIF-2 蛋白将结合的 GTP 水解成 GDP，引起该蛋白质构象发生改变并将从核糖体小亚基上释放下来。然后核糖体大亚基与小亚基结合，形成一个完整的核糖体，并由它开始蛋白质的合成（图 6-71）。

因为 eIF-2 与 GDP 结合非常牢固，所以需要一个鸟嘌呤核苷酸交换因子（图 15-54），命名为 eIF-2B，来促使 GDP 释放，这样一个新的 GTP 分子就能结合上来，eIF-2 就能重新得到利用[图7-101(A)]。当 eIF-2 被磷酸化时，它的再利用就受到抑制——磷酸化的 eIF-2 与 eIF-2B 结合得异常牢固，从而使 eIF-2B 失活。细胞内存在的 eIF-2 比 eIF-2B 多，即便一部分磷酸化的 eIF-2 就能捕获几乎全部的 eIF-2B。这防止了非磷酸化 eIF-2 的再利用并大大减慢了蛋白质的合成[图7-101(B)]。

作为允许它们进入非增殖的，静止期（G_0）的机制的一部分，在哺乳动物细胞中对活性 eIF-2 水平的调节特别重要，在此时期整个蛋白质合成速率比增殖细胞中的降低了大约 1/5（讨论见第 17 章）。

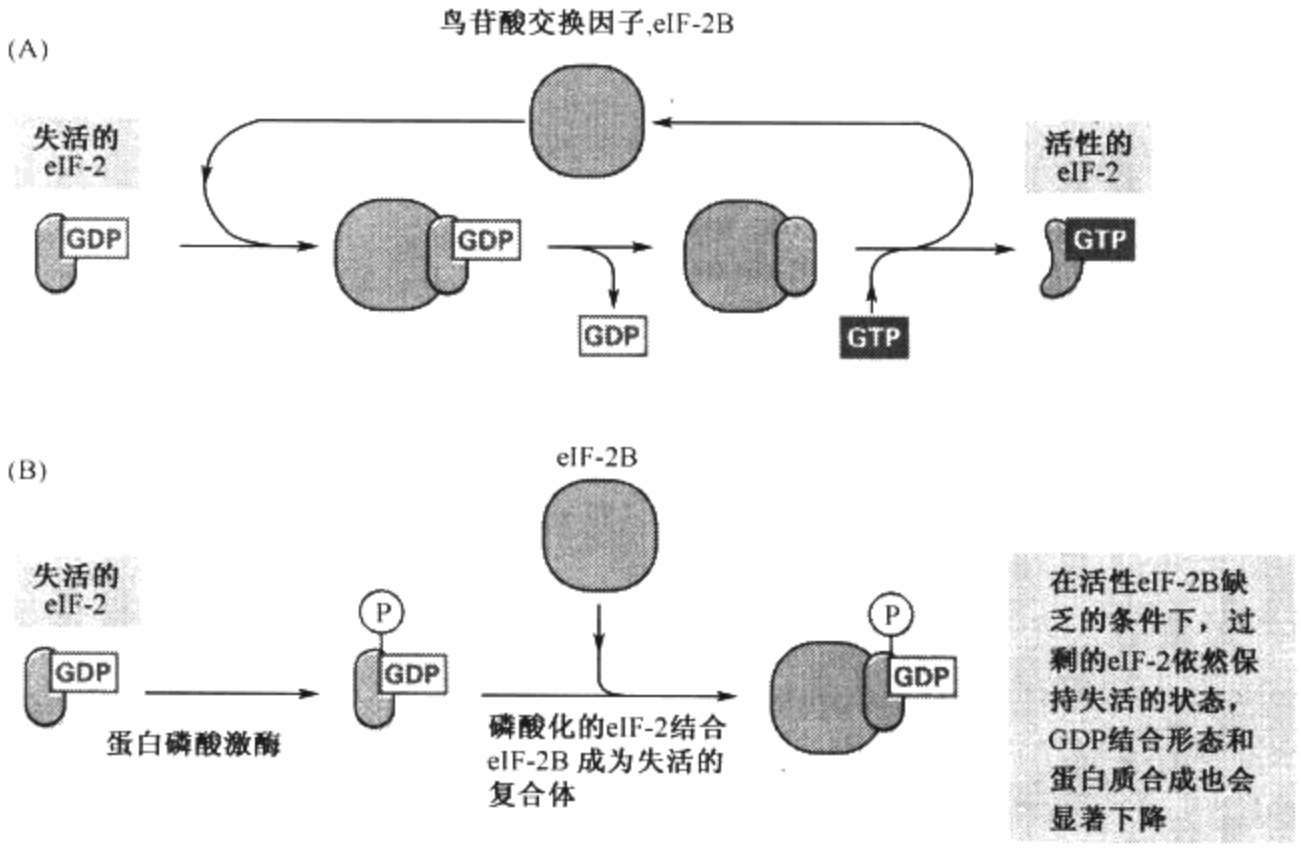


图 7-101　eIF-2 循环

（A）使用过的 eIF-2 由一种鸟嘌呤核苷酸交换因子（eIF-2B）循环利用。（B）eIF-2 磷酸化与 eIF-2B 共同调控蛋白质合成速率

在翻译起始位点上游 AUG 密码子上的起始能调节真核生物的翻译起始过程

第 6 章我们看到，真核生物的翻译通常开始于 mRNA 5′端下游第一个 AUG 位置，因为它是扫描的核糖体小亚基遇到的第一个 AUG。但是紧挨着 AUG 的核苷酸也影响了翻译起始的效率。如果识别位点很少，扫描核糖体亚基将忽略 mRNA 中第一个 AUG 密码子并跳到第二个或第三个 AUG 上。这个现象称作“渗漏扫描”，是经常使用的策略，以便从同一 mRNA 生成两个或多个密切相关的蛋白质，它们只是在氨基末端存在差别。例如，渗漏扫描使有些基因能生成相同的蛋白质，在其氨基端有或没有附着的信号序列，这样该蛋白质就引导至细胞内两个不同的位置。在有些例子中，细胞能调控渗漏扫描生成的蛋白质异型体的相对丰度，例如，起始因子 eIF-4F 的丰度细胞特异性地增加会优先利用靠近 mRNA 5′端的 AUG。

真核生物中发现的另一类调控利用了位于 mRNA 5′端和基因起始处之间一个或多个短可读框，通常这些上游可读框（uORF）编码的氨基酸序列无关紧要，仅纯粹提供了调节功能。mRNA 分子上存在的 uORF 一般会降低下游基因的翻译，这是通过捕获扫描核糖体起始复合体并促使核糖体翻译 uORF，并在核糖体到达蛋白质编码序列之前与 mRNA 解离实现的。

当某个通用翻译因子，例如，eIF-2（上面讨论）的活性下降时，有人可能会想到所有 mRNA 的翻译都会等量地下降。然而，事实恰恰相反，eIF-2 的磷酸化具有选择性的影响，甚至增加含有 uORF 的特定 mRNA 的翻译。例如，这可以使酵母细胞能够适应特定营养物质缺少导致的饥饿，通过关闭所有蛋白质的合成，那些合成该营养物所需的蛋白质除外。作用于一种特定酵母 mRNA 的这种机制的细节已经弄清楚了，该 mRNA 编码一个称作 GCN4 的蛋白质，它是

一个基因调节蛋白，是活化许多编码对氨基酸合成很重要的蛋白质基因所必需的。

GCN4 mRNA 含有 4 个短 uORF，它们负责选择性地增加 GCN4 的翻译，以应答氨基酸饥饿引起的 eIF-2 磷酸化。这个增加 GCN4 翻译的机制很复杂。扼要地讲，核糖体亚基沿 mRNA 移动，会遇到每个 uORF，但只翻译它们的一部分；如果第 4 个 uORF 翻译了，这在未挨饿细胞中是事实，核糖体就在此 uORF 末端解离，GCN4 的翻译就无效了。eIF-2 活性的下降使扫描核糖体在获得起始翻译的能力之前更可能通过第 4 个 uORF。然后，这样的一个核糖体有效地起始 GCN4 序列上的翻译，从而合成促进细胞内氨基酸合成的蛋白质。

内部核糖体进入位点提供了翻译调控的机会

虽然将近 90%的真核生物 mRNA 的翻译是从 5′帽子下游第一个 AUG 开始的，但是我们将在最后一部分看到，某些 AUG 在扫描过程中能够被略过。这一部分，我们仍然讨论另一个细胞从远离 mRNA 5′端位置起始翻译的方式。在这些情况下，翻译直接在特殊 RNA 序列上起始，它们每一个都称作**内部核糖体进入位点**（IRES）。mRNA 中许多不同位置都可以出现 IRES。在某些少见的情况下，相同真核生物 mRNA 中携带了串联的两个不同蛋白质编码序列：第一个序列的翻译使用通常的扫描机制，而第二个通过 IRES 翻译。IRES 通常有几百个核苷酸长，并折叠成特定的结构，这些结构结合许多，但不是全部，用于起始正常帽子依赖翻译的蛋白质（图 7-102）。实际上，不同的 IRES 需要一小组不同的起始因子。然而，它们全部都不需要 5′帽子结构和识别帽子的翻译起始因子 eIF-4E。

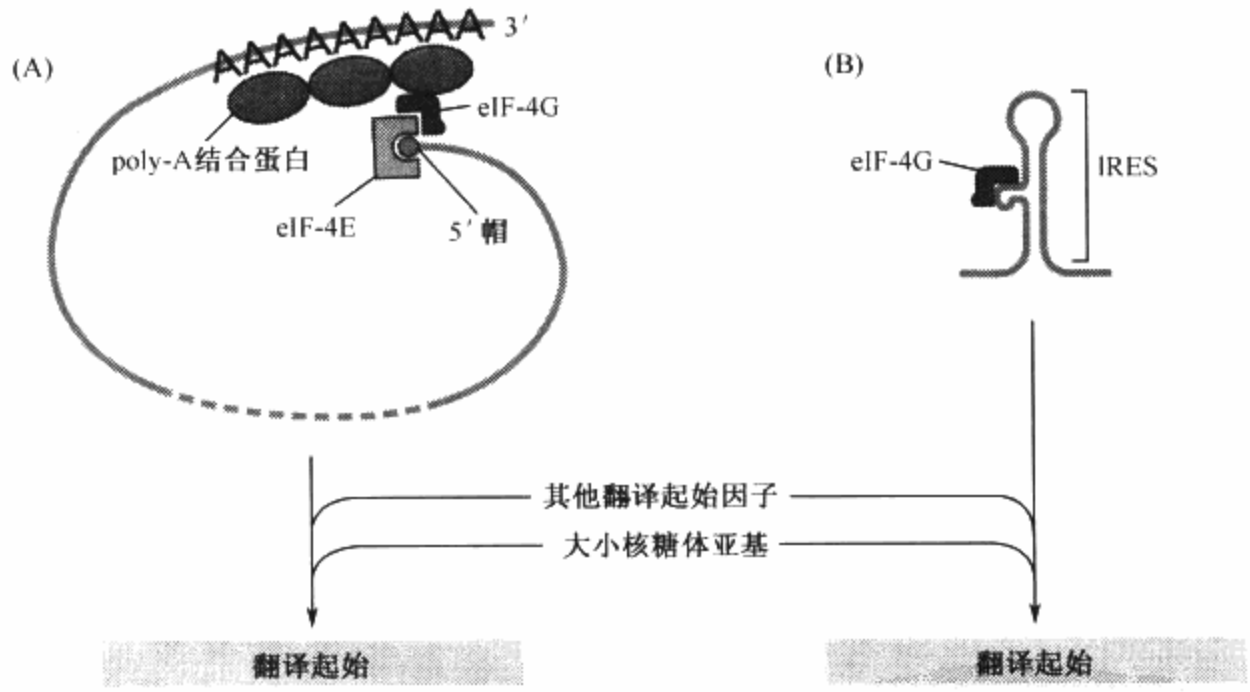


图 7-102　两种翻译起始机制

（A）帽依赖机制需要一组起始因子，5′帽和多聚 A 尾的存在刺激了它们在 mRNA 上的组装（图 6-71）。（B）IRES 依赖机制只需要常规翻译起始因子中的一小部分，它们直接在折叠的 IRES 上组装（摘自 A. Sachs，*Cell* 101：243～245，2000）

IRES 最先在某些哺乳动物病毒中发现，在这里它们提供一种精巧的方式让病毒占用宿主细胞的翻译装置。感染后，这些病毒合成一种蛋白酶（编码在病毒基因组中），它切割细胞翻译因子 eIF-4G 因而使其不能与帽子结合复合物 eIF-4E 结合。这将大部分宿主细胞翻译关闭掉并有效地将翻译装置改换到许多病毒 mRNA 上都存在的 IRES 序列。截短的 eIF-4G 仍然能起始这些内部位点的翻译，甚至可能刺激了某些含有 IRES 的病毒 mRNA 的翻译。

细胞 mRNA 也发生 IRES 介导翻译的选择性激活。例如，当真核细胞进入细胞周期的 M 期时，翻译的整体速度下降到间期细胞中的 25%左右。这种降低主要是因为帽子结合复合物——eIF-4E 发生依赖于细胞周期的去磷酸化从而降低了它对 5′帽子的亲和力。然而含有 IRES 的 mRNA 对这种效应产生免疫。因而细胞进入 M 期后，它们的相对翻译速率会增加。

最后，当哺乳动物细胞进入程序性细胞死亡途径时（讨论见第 17 章），eIF-4G 被切割，随后再次发生整体性的翻译减少。有些对细胞死亡调控很关键的蛋白质从含有 IRES 的 mRNA翻译出来，并且它们持续被合成。似乎对于细胞而言 IRES 机制的主要优势之一就是，它使经过选择的 mRNA 以高速率翻译，尽管细胞起始蛋白质合成的能力整体上下降了。

mRNA 稳定性的改变能调控基因表达

细菌细胞内大多数 mRNA 都很不稳定，半衰期大约是 3min。以 3′→5′方向进行降解的外切核酸酶通常负责这些 mRNA 的迅速降解。因为细菌的 mRNA 既迅速地合成，又快速地被降解，所以它能迅速地适应环境的变化。

真核细胞中 mRNA 要更加稳定一些。有一些半衰期超过 10h，如编码 β-珠蛋白的 mRNA。然而许多半衰期也只有 30min 或者更短。这些不稳定 mRNA 通常编码调节蛋白，如生长因子和基因调控蛋白，它们在细胞内的合成速率需要迅速地发生变化。

真核生物有两个主要的降解 mRNA 的途径，并且每个 mRNA 分子中的序列决定了所用途径以及降解的动力学。最常见的途径涉及 poly A 尾的逐渐变短。第 6 章我们看到 mRNA 分子的帽化和多聚腺苷化在细胞核内进行。一旦进入细胞质，poly A 尾（平均长度大约为 200 个 A）逐渐被一种以 3′→5′方向消化尾巴的外切核酸酶作用而变短。一旦尾巴缩短过程到达一个关键时期（约剩下 30 个 A），就移走 5′帽子（此过程称为脱帽化），而 RNA 就迅速地被降解掉[图7-103(A)]。

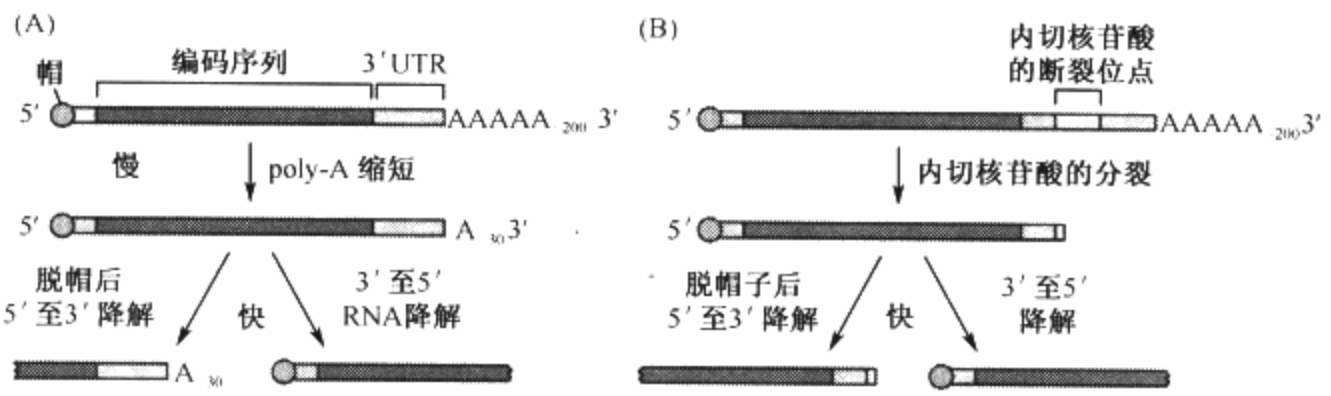


图 7-103　真核 mRNA 衰变的两种机制

(A) 脱腺苷化依赖的衰变。大部分真核 mRNA 经过这种途径降解。诱导衰变的多聚 A 尾长度临界值可能对应于多聚 A 结合蛋白的丢失（图 6-40）。如图 7-104 所示，脱腺苷化酶与 3′多聚 A 尾和 5′帽子相连，这种安排可能协调了脱帽过程和多聚 A 缩短过程。虽然所显示的 5′→ 3′降解和 3′→5′降解发生在单独的 RNA 分子上，但这两个过程可以在同一分子上进行。(B) 独立于脱腺苷化的衰变，还不能确定是否脱帽化紧随 mRNA 的内切核酸酶解切割之后发生（摘自 C. A. Beelman and R. Parker, *Cell* 81: 179～183, 1995）

几乎所有的 mRNA 都要发生 poly A 尾缩短、脱帽化和最终的降解过程，但不同物种的 mRNA之间这个过程发生的速率不同。行使缩尾功能的蛋白质直接与催化翻译的装置竞争；因此，任何影响 mRNA 翻译效率的因子都将倾向于对降解有相反的作用（图 7-104）。此外，许多 mRNA 的 3′UTR 中携带增加或降低 poly A 尾缩短速度的特殊蛋白质结合序列。例如，许多不稳定的 mRNA 含有 AU 序列片段，它大大提高了缩短速率。

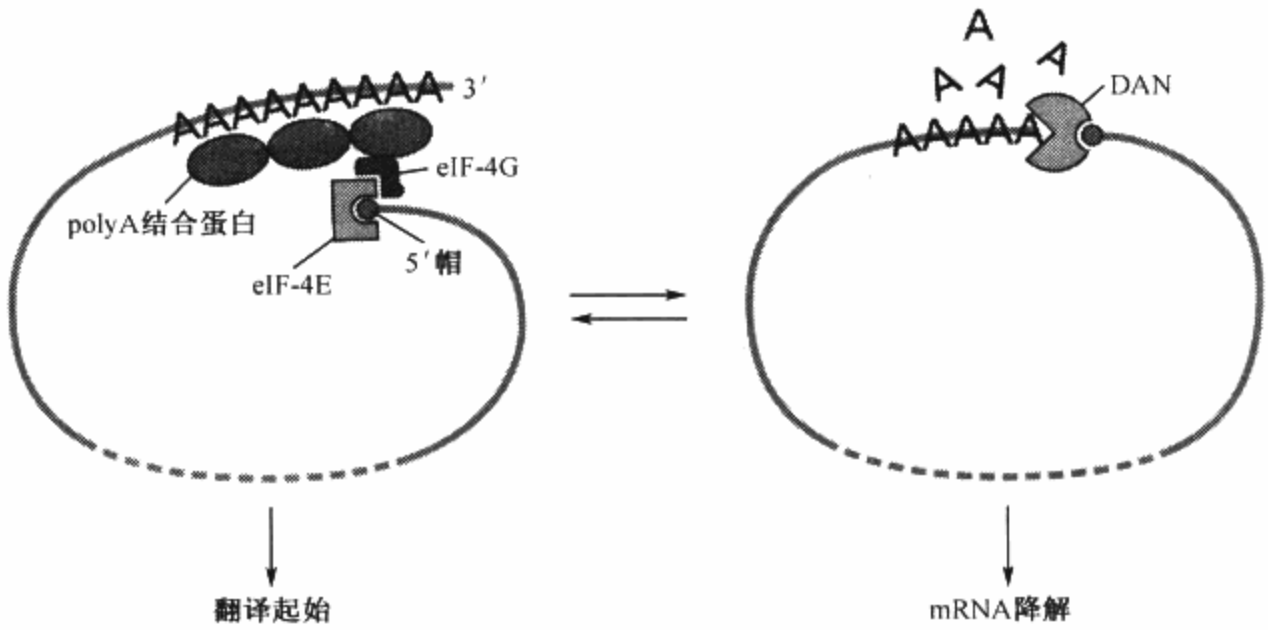


图 7-104　mRNA 翻译和 mRNA 衰变之间的竞争

翻译起始和脱腺苷化依赖的 mRNA 衰变都利用了 mRNA 两个相同的特征——5′帽子和 3′ polyA 位点。以 3′→5′方向缩短 polyA 尾的酶（称为 DAN）与 5′帽子结合。如第 6 章所述（图 6-71），翻译起始装置也与 5′帽子和 poly A 尾结合（摘自 M. Gao et al.,*Mol. Cell* 5：479～488，2000）

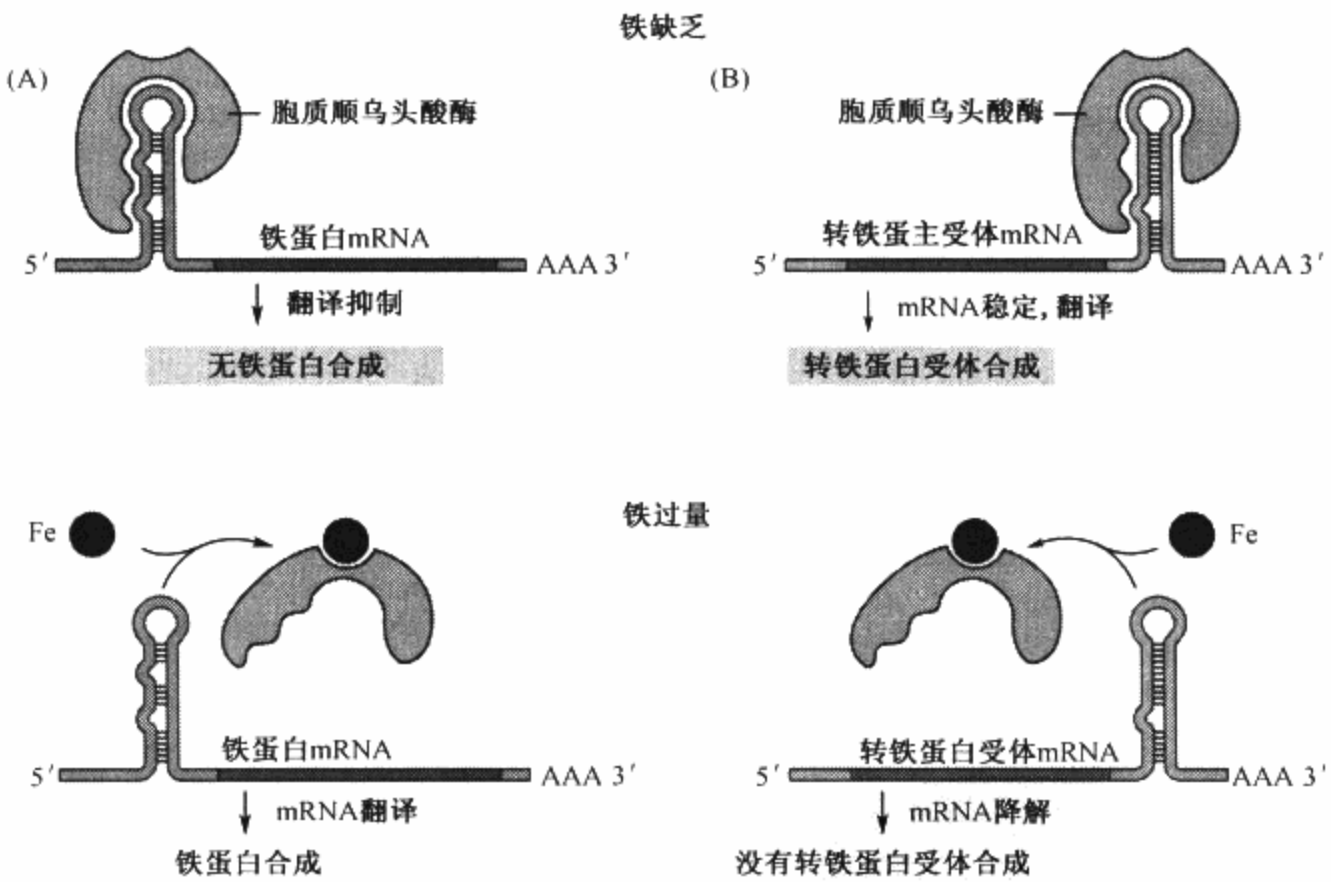


图 7-105　两种由铁介导的翻译后调控

为了应答细胞质中铁浓度的升高，细胞会增加铁蛋白的合成，以便结合过量的铁（A）并减少转铁蛋白的合成，以便使跨质膜输入铁减少（B）。这两种应答都由同一铁应答调节蛋白——顺乌头酸酶介导，它识别编码铁蛋白和运铁蛋白的 mRNA 中的茎环结构的共同特征。当顺乌头酸酶结合上铁以后就从 mRNA 上解离。但由于转铁蛋白受体和铁蛋白受不同机制的调控，所以它们的水平以相反的方向对应于铁浓度，即便它们受相同的铁响应调节蛋白的调控。顺乌头酸酶结合铁蛋白受体 mRNA 的 5′UTR 阻断了翻译起始；它结合铁蛋白受体 mRNA 3′UTR 则阻断一个内切核酸酶位点，因而稳定了 mRNA（图 7-103）（摘自 M. W. Hentze et al.,*Science* 238：1570～1573，1987 and J. L. casey et al.,*Science* 240：924～928，1988）

另一种 mRNA 降解途径开始于特定内切核酸酶的作用，它作用一步只是将 poly A 尾从 mRNA 的其余部分切割下来（图 7-103）。以这种方式降解的 mRNA 通常在它们的 3′UTR 中携带了特殊的核苷酸序列作为内切核酸酶的识别序列。

mRNA 的稳定性在应答细胞外的信号时能发生改变。例如，向细胞加入铁会降低编码与铁转运蛋白结合的受体蛋白 mRNA 的稳定性，从而减少这个受体蛋白的合成。有意思的是，这个效应由铁敏感 RNA 结合蛋白——顺乌头酸酶介导，我们在前面也讨论过，这个蛋白质也调控铁蛋白 mRNA 的翻译。顺乌头酸酶能结合转铁蛋白受体 mRNA 的 3′UTR，并通过阻断该 mRNA 内切核酸酶解切割，使受体合成量增加。加入铁后，顺乌头酸酶从 mRNA 释放，从而降低了 mRNA 的稳定性（图 7-105）。

细胞质内 poly A 加成能调控翻译

RNA 分子最初的多聚腺苷化（讨论见第 6 章）在细胞核中进行，显然这对于几乎所有真核生物 mRNA 前体都是自发进行的。我们刚看到了，大多数 mRNA 的 poly A 尾在细胞质中逐渐缩短，最终 RNA 被降解。在此情况下，特定 mRNA 的 poly A 尾在细胞质中会延长，这种机制提供了又一种翻译调控形式。

成熟过程中的卵母细胞和卵提供了最显著的例子。在这些巨大的细胞中许多常规的 mRNA降解途径似乎失去了功能，以至于这些细胞在为受精做准备时能积累大量的 mRNA。许多 mRNA 储存在细胞质中，其 3′端只有 10～30 个 A，以这种形式存在时，它们是不能被翻译的。在卵母细胞成熟期间和受精后的特定时刻，需要这些 mRNA 编码的蛋白质时，poly A就加到经过选择的 mRNA 上，从而大大刺激它们翻译的起始。

无义介导 mRNA 降解在真核生物中用作一种 mRNA 监视系统

第 6 章我们看到，真核生物中 mRNA 的合成需要经过一系列设计复杂的步骤。只有当 mRNA合成的所有这些步骤都完成了，mRNA 才从细胞核输出，到达细胞质翻译成蛋白质。如果这些步骤中任何一步出现错误，RNA（与切除的内含子一起）最终就会在细胞核被外核体降解掉，后者是一种含有至少 10 种 3′→5′RNA 外切核酸酶的巨大蛋白质复合体。真核细胞还具有另外一种机制，称为**无义介导 mRNA 降解**，在异常 mRNA 能开始有效地翻译成蛋白质之前将它们清除。当发现含有错位的框内终止密码子（UAA、UAG 或 UGA）的 mRNA 会被迅速降解时，发现了这种机制的存在。这些终止密码子可能因突变或不完全剪接而产生：两种情况下都观察到这个现象。这个 mRNA 监视系统因而防止了合成截短的异常蛋白质，我们已经看到，这类蛋白质对细胞非常危险。但是细胞是如何识别那些潜在的有害 mRNA 呢?

在脊椎动物中，无义介导 mRNA 降解系统感受到的 mRNA 的关键特征，是第一个框内终止密码子和 RNA 剪接形成的外显子-外显子边界之间的空间关系。如果这个终止密码子位于所有外显子-外显子边界的下游（3′端），mRNA 就不会被无义介导降解；另一方面，如果有一个终止密码子位于一个外显子边界上游（5′），mRNA 就被降解掉。进行翻译的核糖体，以及其他的监视蛋白质会为每个 mRNA 评估这种关系。到底这是如何实现的，具体还不清楚，但容易理解为什么必须有核糖体的参与：只有框内终止密码子引起无义介导的降解，并且正是核糖体与 mRNA 之间的关系决定了可读框。根据一个模型（图 7-106），核内蛋白质结合并因而标记了 RNA 剪接后的外显子-外显子边界。当 mRNA 离开细胞核以后，它保留在核周围，当翻译开始以后由一组其他的监视蛋白质连接起来。根据这个观点，mRNA 分子第一轮翻译将只用于评估这条 mRNA 是否适用于进一步的翻译。如果 mRNA 通过了这个测试，

当此 mRNA 被释放并在细胞质扩散时就会迫不及待地进行翻译。

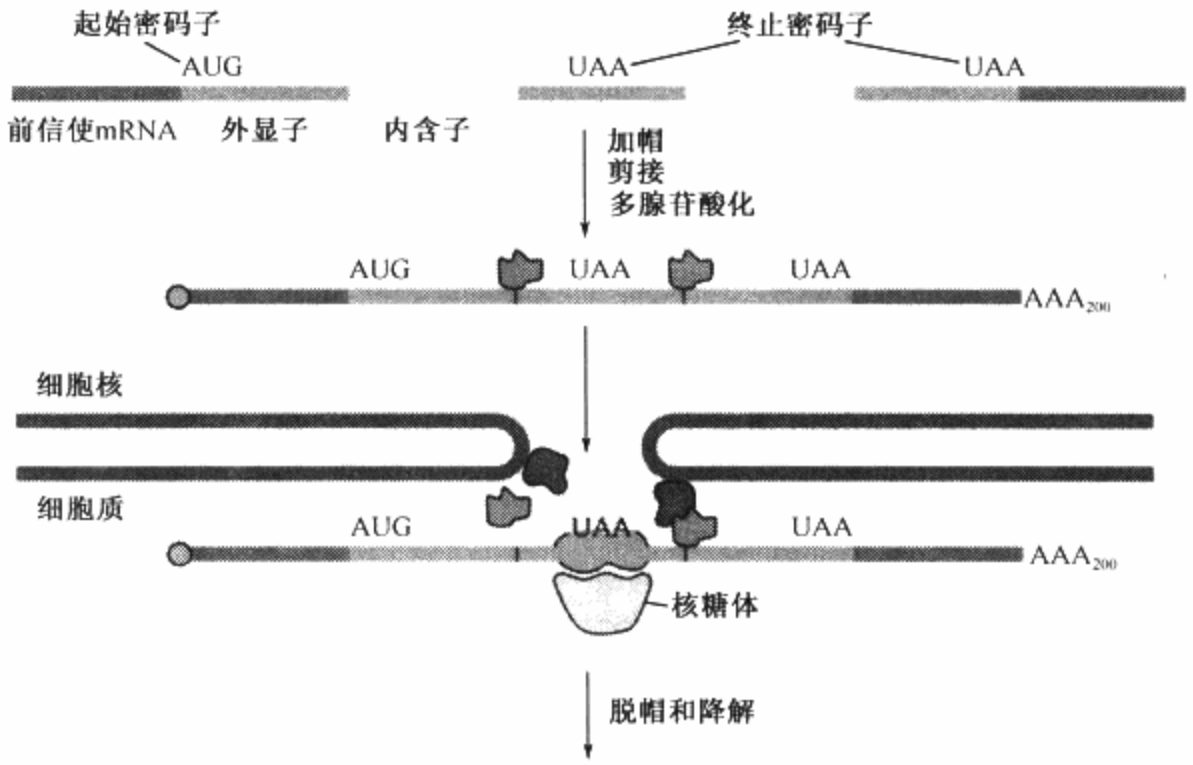


图 7-106　无义介导的 mRNA 降解模型

根据该模型，核蛋白（橙色）标记剪接后的 mRNA 分子上的外显子-外显子边界。这些蛋白质据认为随剪接反应进行组装，可能也参与了成熟 mRNA 从细胞核向外转运（图 6-40）。成熟 mRNA 由核内输出并停留在核膜附近，在这里由核糖体在其他监控蛋白帮助下（绿色）进行一轮检验性的翻译（深绿色）。如果在到达最后的外显子-外显子边界前遇到框内终止密码子，mRNA 就会进行无义介导 mRNA 降解。如果没有，mRNA 就从核被膜释放（可能是由于核糖体取代了外显子-外显子标记蛋白的位置），并能在细胞质中自由地进行多轮翻译。根据显示的模型，这轮检验性翻译在核外进行，然而它也可能发生在核内，在 mRNA 输出之前（摘自 J. Lykke-Anderson et al.,*Cell* 103：1121～1131，2000）

无义介导的降解在进化中可能发挥着特别重要的作用，使真核细胞能够更加容易地通过 DNA 重排、突变或选择性剪接模式探索新基因，从而仅仅选择出那些合成出长度完整的蛋白质 mRNA。无义介导降解在发育的免疫系统细胞中也很重要，在这些细胞中发生的更广泛的重排（图 24-37）常常生成提前终止的密码子。这种重排的基因产生的 mRNA 被这个监视系统降解掉，因而防止截短的蛋白质对细胞产生有害作用。

细胞利用 RNA 干扰来沉默基因表达

真核细胞利用一种特殊类型的 RNA 降解作为一种防御机制来破坏外来 RNA 分子，尤其是那些借助它们在细胞内以双链形式存在才得以鉴定出来的分子：这种机制称为 **RNA 干扰（RNAi）**，它在大量的生物中都有发现，包括单细胞真菌、植物、蠕虫、小鼠，可能还有人类，表明它在进化上是一种古老的防御机制。在植物中，RNA 干扰保护细胞不受 RNA 病毒的侵害。在其他类型的生物中，认为它防止了经由 RNA 中间产物进行复制的转座元件的增殖（图 5-76）。许多转座元件和植物病毒在它们的生命周期中都合成双链 RNA，至少是暂时地合成。RNAi 不仅有助于控制这种侵扰，而且为科学家提供一种强大的实验技术来关闭单个细胞基因的表达（图 8-65）。

游离双链 RNA 的存在通过吸引含有一个 RNA 核酸酶和一个 RNA 解螺旋的蛋白质复合体来引发 RNAi。这个蛋白质复合体将双链 RNA 切割成小片段（将近 23 个核苷酸对），它们

仍然和酶保持连接。这些结合的 RNA 片段接着引导该酶复合体到其他的具有互补核苷酸序列的 RNA 分子上，并由这个酶将这些分子降解掉。这些其他的分子可以是单链或者是双链的（只要它们有一条互补的链）。通过这种方式，科学家能利用实验方法引入一条双链 RNA 分子将特定的细胞 mRNA 失活（图 7-107）。

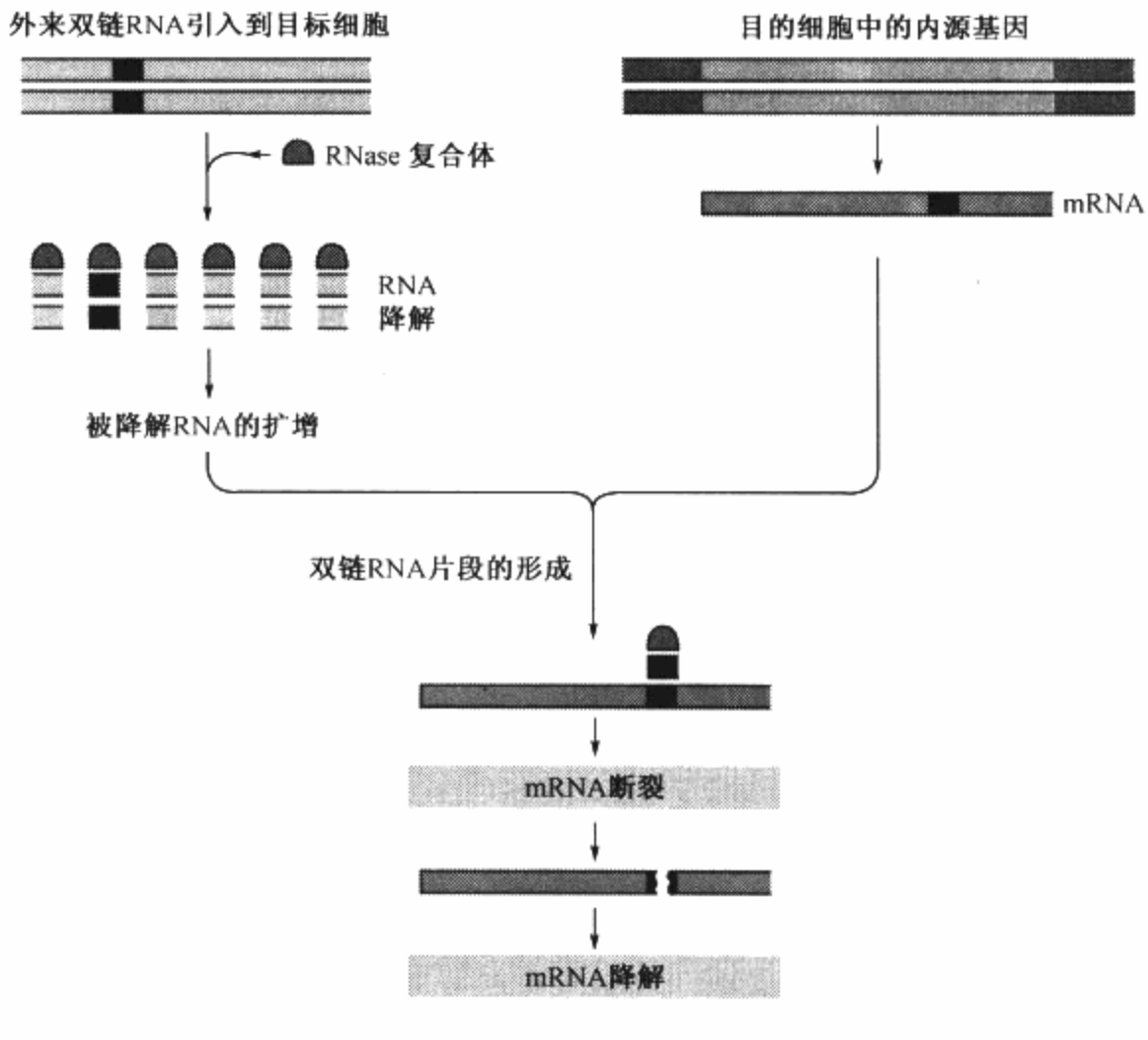


图 7-107　RNA 干扰机制

左边显示的是外来双链 RNA 分子的命运，它们由一个巨大蛋白质复合体中的 RNase 酶识别，并降解成长度大约 23 个核苷酸对的小片段。这些片段有时由依赖于 RNA 的 RNA 聚合酶扩增，并且在这种情况下，能有效地传递到子细胞中。如果外来 RNA 核苷酸序列与细胞基因的类似（图的右边），那么该基因成熟的 mRNA 也会通过图中所示的途径被降解掉。通过这种方式，可以利用实验方法向细胞引入与基因核苷酸序列配对的双链 RNA 来关闭掉细胞基因的表达。RNA 干扰也需要 ATP 水解和 RNA 解螺旋酶，可能是为了产生能与其他 RNA 分子形成碱基对的单链 RNA 分子

这个酶复合体每次切割一个新 RNA 的时候，就与一个短 DNA 分子一起再生，这样最初的双链 RNA 分子能发挥催化作用破坏许多互补 RNA。此外，短双链 RNA 切割产物自身能被其他的细胞酶复制，从而提供了甚至更加扩大的 RNA 干扰活性（图 7-107）。这种放大作用确保了一旦起始，RNA 干扰就能持续下去，即便在起始的双链 RNA 已全部被降解或者被稀释时也如此。例如，它允许后代细胞继续携带在亲代细胞中激发的 RNA 干扰活性。此外，RNA 干扰活性能因 RNA 片段从一个细胞向另一个细胞的转移而得以扩展。这在植物中尤其重要（它们的细胞之间存在精细的连接通道，讨论见第 19 章），因为它使整个植株仅仅在少数细胞被感染后就能对 RNA 病毒产生抵抗力。

小结

从 RNA 到蛋白质途径中允许许多步骤都受到细胞的调节，以便调控基因的表达。大部分基因据认为都在多种水平上受到调控，虽然转录起始的调节（转录调控）常常占主要地位。

然而，有些基因以恒定的水平转录并只由转录后调节过程打开或者关闭。这些过程包括①RNA 转录提前终止导致其转录物的弱化；②选择性剪接位点的选择；③切割或 poly A 加成对形成 3′端的调控；④RNA 编辑；⑤从细胞核向细胞质转运的调控；⑥mRNA 在细胞特定部位的定位；⑦翻译起始的调控；⑧可调节的 RNA 降解。大部分这些调控过程需要识别受调控的 RNA 分子中特定的序列或结构。这种识别由调节蛋白或调节 RNA 分子来完成。

基因组如何进化

本章以及前面的 3 章，我们讨论了基因的结构，它们在染色体上的排列方式，将遗传信息转变为功能蛋白和 RNA 分子的复杂的细胞装置，以及细胞调控基因表达的许多途径。这一部分，我们讨论基因和基因组长期以来经过进化，产生我们地球上多种多样的生命形式的一些方式。基因组测序根本性地改变了我们有关分子进化过程的观点，揭示了大量令人惊奇有关生物和进化机制中家族关系的信息。

在广泛不同的生命体中发现具有类似功能的基因可能并不让人奇怪。但过去 20 年来最重要的事实却是发现实际上许多基因的序列都非常保守，以至于跨过巨大的种系发生距离仍然能常常发现同源基因，也就是，由于共有相同的祖先导致核苷酸序列相似的基因。例如，在线虫、果蝇、酵母，甚至细菌这些生物中能找到许多人类基因显而易见的同系物。

第 3 章讨论过，并将在第 8 章再次讨论到，识别序列同源性已成为推理基因和蛋白质功能的主要方式。尽管找到这样一种同系物并不能保证在功能上相似，但已证明它是一种很好的线索。因此，通过比较一个还没有生物化学和遗传信息可利用的人类基因序列和其他生物中经过广泛研究的基因序列，有可能估计出这个基因的功能。

基因序列常常比整个基因组结构保守性更强。第 4 章讨论过，基因组的组织特征，例如，基因组大小、染色体数目、基因在染色体上的排列顺序、内含子的丰度和大小、重复 DNA 的数量在不同生物中发现差别很大，基因的实际数目也是如此。

基因数目与生物的表型复杂度只是大致上相关。例如，目前估计面包酵母的基因数是 6000 个，线虫的基因数是 18 000 个，果蝇的基因数是 13 000 个，人类基因数是 30 000 个（表 1-1）。我们将很快看到，基因数目随生物复杂度增加而增加大多涉及紧密相关的基因家族的扩增，这个事实将基因重复和趋异确立为主要的进化过程。事实上，可能所有现今基因都是存在于早期生命形式中的一些古老基因的后裔——经由复制、趋异和基因片段重新搭配的过程。

拷贝和维持 DNA 的正常机制失灵引起基因组改变

除少数例外，细胞不具备改变它们基因组结构的特殊机制：进化而是依赖于偶然事件和错误。所发生的大部分遗传改变只是由于基因组拷贝或损伤后修复的正常机制失灵而产生，尽管可移动 DNA 元件也起着重要作用。第 5 章我们曾讨论过，维持 DNA 序列的机制相当精确，但是它们并不是完美无瑕的。例如，由于使 DNA 序列能够高度忠实地遗传下去的复杂 DNA 复制和 DNA 修复机制的存在，每 20 万年大约每 1000 个核苷酸对中只有一个发生随机改变。即使如此，在一个具有 10 000 个体的群体中，100 万年时间内每个可能的核苷酸替换都检验了约 50 次——对于物种进化来说，100 万年只是一段很短的时间而已。

DNA 复制、DNA 重组或 DNA 修复中出现错误或者能引起 DNA 序列的简单改变，例如，一种碱基对替代了另一种碱基对，或者引起大规模的基因重排，如删除、重复、倒位和

DNA 从一个染色体向另一个染色体的转位。有人认为这些错误发生的速率本身已被进化过程定型，以提供一个基因组稳定和变化之间可接受的平衡。

除复制和修复机制失灵以外，第 5 章描述过的各种可移动 DNA 元件也是产生遗传变化的重要来源。尤其是，作为占据基因组的寄生 DNA 序列，转座 DNA 元件（转座子），发挥主要作用并能在基因组中散布。在此过程中，它们常常干扰所处基因的功能或改变它们的调控；有时通过转座序列和现有基因融合，它们甚至共同生成新的基因。表 5-3 提供了 3 类主要的转座子实例。经历长期的进化以后，这些转座子已经深远地影响了基因组的结构。两种物种间基因组序列的差异与它们分别进化的时间长短呈比例。

两个物种基因组序列的差别与它们独自进化以来经历的时间长短呈比例

现存物种基因组之间的差别是经历至少 30 多亿年后积累而成的。虽然缺少长期以来发生改变的直接记录，可是我们能够对当今生物基因组进化详细的比较后，重建出基因组进化的过程。

比较基因组学的基本工具是系统树。一个简单的例子就是描述人类从巨猿趋异进化的系统树（图 7-108）。这个树的主要支持证据来源于基因和蛋白质序列的比较。例如，人类基因或蛋白质的编码序列和巨猿之间的比较典型地揭示出人类和黑猩猩之间差异最小而人类和猩猩之间差异最大。

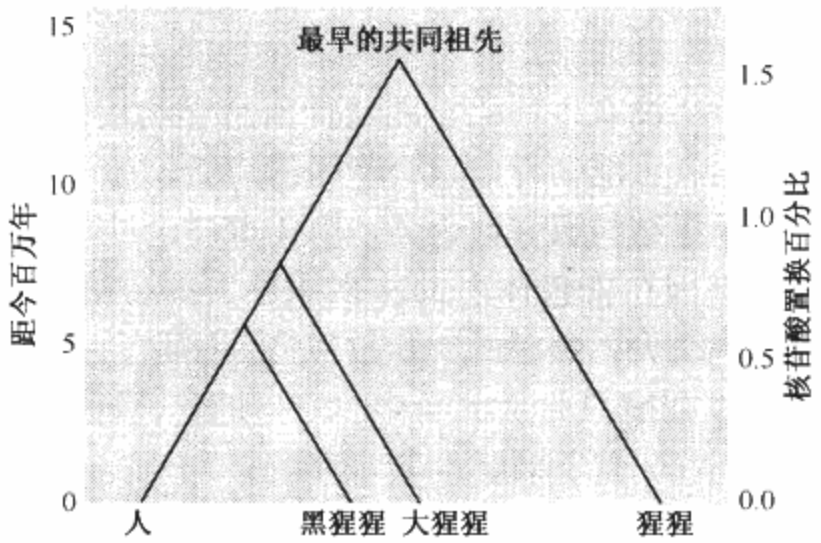


图 7-108　基于核苷酸序列数据的显示人类和巨猿之间关系的系统树

如图所示，据估计，全部 4 个物种基因组序列与它们最近的共同祖先基因组序列差异达 1.5%还多一点。由于变化在两个趋异种系独立产生，所以逐对比较揭示了 2 倍的与最近的共同祖先相比的序列趋异程度。例如，人类-猩猩比较典型地表明序列趋异达 3%多一点，而人类-猩猩比较表明趋异约达 1.2%（摘自 F. -C Chen and W. -H. Li, *Am. J. Hum. Genet.* 68：444～456，2001）

对于密切相关的生物如人类和黑猩猩来说，有可能重建出这两个物种已灭绝的最近共同祖先的基因序列（图 7-109）。人类和黑猩猩之间紧密的相似性主要是由于在这两个趋异的种系可以用来累积突变的时间太短，而不是由于功能限制将这些序列保持相同，这种观点的证据来源，在于观察到在功能上即使核苷酸顺序不受限制的 DNA 序列也几乎一模一样，例如，编码血纤肽的序列或者“同义”密码子（指令相同氨基酸的密码子，图 7-109）的第三个位置。

对于相关性不太密切的生物，如人类和小鼠，基因中发现的序列保守性主要是由于**纯化选择**（也就是说除去携带有干扰重要基因功能的突变个体），而不是由于允许发生突变的时间

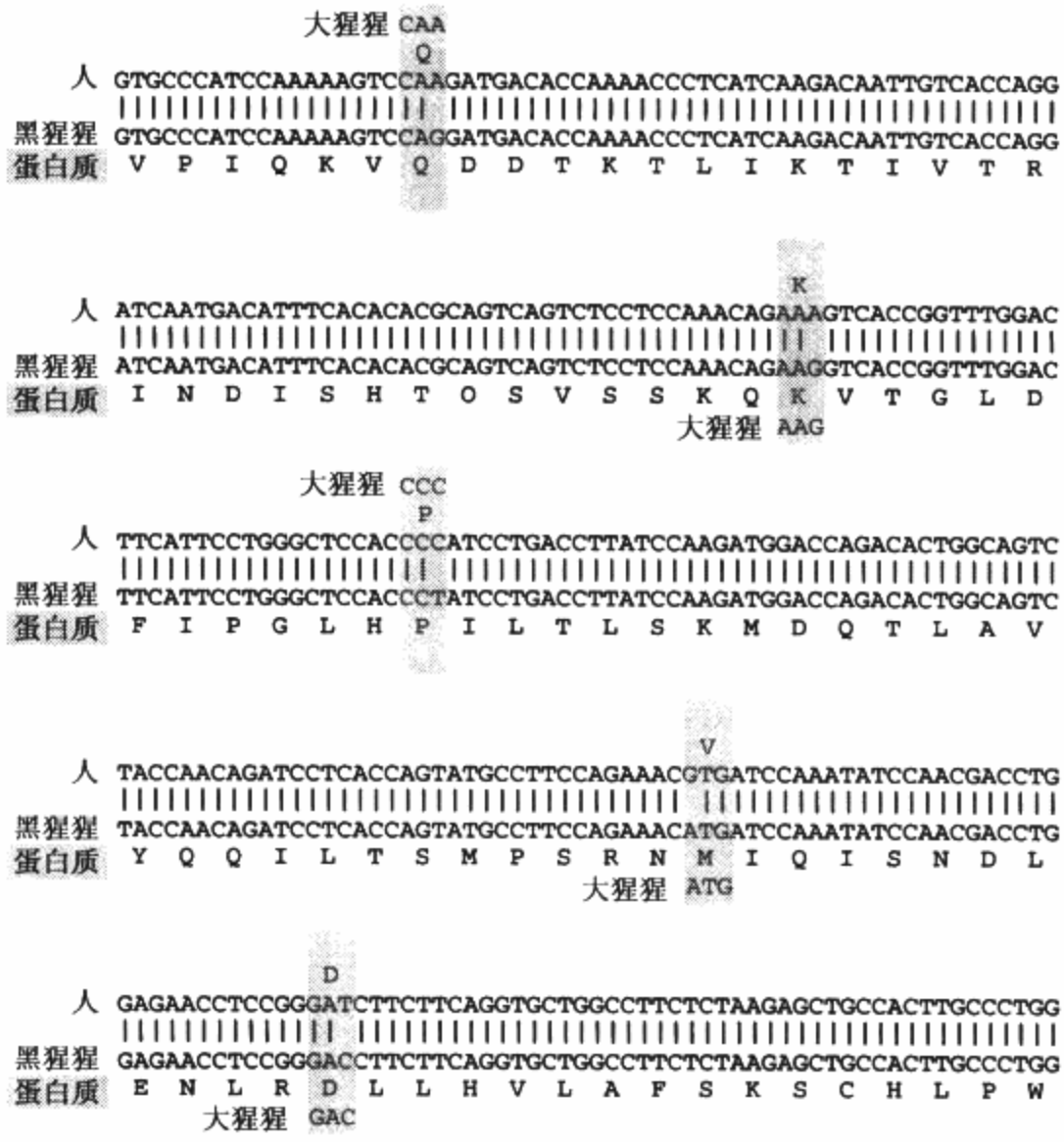


图 7-109　从人类和黑猩猩瘦蛋白基因编码区序列比较中追溯始祖序列

瘦蛋白是一种调节食物摄取和能量利用以应答脂肪储备量的激素。如绿色表示的密码子框所示，这两个序列中只有 5 个密码子不同（总共 441 个核苷酸）。此外，如果核查一下人类和黑猩猩的序列编码的氨基酸，就会发现这 5 个位置中只有一个编码的氨基酸不同。图中也显示了大猩猩中这 5 个不同位点每一个的相应序列。在两种情况下，大猩猩的序列与人类序列一致，而在 3 种情况下与黑猩猩的序列一致。那么最近的共同祖先中瘦蛋白基因的序列是什么呢？当这些序列一致时，一种设法将人类和黑猩猩基因在进化过程中假定发生过的突变数目最小化的进化模型，把最近的共同祖先瘦蛋白序列与人类和黑猩猩序列相同作为前提；如果这些序列不一致，就利用大猩猩的序列作为突破口。为了方便起见，只给出了瘦蛋白编码序列前 300 个核苷酸。剩余的 141 个核苷酸在人类和黑猩猩之间是相同的

不够长。结果，限制用于与保守蛋白质发生高度特异相互作用的 DNA 中，蛋白质编码序列和调控序列常常显著地保守。相比之下，人类和小鼠基因组中大多数 DNA 序列趋异得如此之远以至于常常不可能将它们彼此对比起来。

以分子序列和化石记录的比较作为基础的系统树整合产生了最合适的看待现代生命形式进化的观点。基于发现化石的岩石结构中放射性同位素衰变，化石记录作为鉴定绝对年代的来源，仍然很重要。然而，从化石记录难以建立物种间精确的趋异次数，即使对于那些留下具有不同形态的完好化石的物种来讲也是如此。在一个新的物种数量扩大到足够多以至于留下可以发现的化石记录以前，该种群长期以来可能都很小并在地理上受局限。此外，即使一个化石与目前的物种密切相似，也不能肯定它就是这个物种的祖先，这个化石可能来自于已灭绝的种系，而当前物种的祖先可能仍然不知道。

整合的系统树支持这样的基本观点，就是在特定基因或蛋白质序列中以恒定的速率发生着变化，至少在世代时间和整体生物学特征彼此非常相似的物种的种系中是这样。分子钟假

说就是指序列改变的这种明显的恒定速率。第 5 章描述过，在没有经历纯化选择的序列中分子钟运行得最迅速，这些序列包括，基因间区域、缺少剪接或调控信号的内含子部分，以及因突变导致不可逆失活的基因（所谓的假基因）。在遭受很强的功能限制的序列中分子钟运行得最慢，例如，和大量其他蛋白质特异地相互作用并因而结构高度受限的蛋白质，如肌动蛋白的氨基酸序列（图 16-15）。

由于分子钟运行的速率由突变速率和特定序列纯化选择的数量决定，所以细胞内由不同系统复制和修复的基因需要不同的标定刻度。最明显的是，由于线粒体的高突变性，以功能上不受限制的线粒体 DNA 序列为基础的分子钟比以功能上不受限制的核内序列为基础的分子钟的运行要快得多。

分子钟比化石记录具有更精细的时间分辨率，比起经典的以不同物种形态和发育的比较为基础的系统树构建方法，也是一个更可靠的构建系统树详细结构的指导方法。例如，直到 20 世纪 80 年代积累了足够分子序列数据因而得以产生图 7-108 中所示的系统树时，巨猿和人类种系之间精确的关系才获得解决。

人类和黑猩猩的染色体非常相似

我们刚才看到不同物种同源基因之间序列相似性的程度由这两个物种共有祖先之后经历的时间来决定。相同的原理也适用于规模更大的基因组结构改变。

人类和黑猩猩基因组，它们独自进化的历史达 500 万年，在整体组织方面仍然几乎完全一样。不仅人类和黑猩猩看起来拥有本质上相同的一套 30 000 个基因，而且这些基因在两个物种的染色体上以几乎相同的方式排列（图 4-57）。唯一的重要差别就是人类 2 号染色体由两条染色体融合而成，而它们在黑猩猩、大猩猩和猩猩中都是单独存在。

即便是转座子活性产生的基因组大规模重塑对于 500 万年规模的人类-黑猩猩趋异也只有很小的影响。例如，两者基因组中都存在的反转座子 Alu 家族的 100 万个拷贝中，超过 99%都处于对应的位置上。这个事实表明在人类和黑猩猩种系趋异之前，我们基因组中大部分 Alu 序列经历了重复和转座。然而，Alu 家族仍然在活跃地转座。因此，目前已观察到一些例子，其中新的 Alu 插入导致人类遗传疾病；这些情况涉及这个 DNA 转座到患者父母基因组中未被占据的一个位点。从更一般意义上讲，存在一类“人类特异”的 Alu 序列，它们占据人类基因组中某些在黑猩猩中未被占据的位点。由于似乎缺少精确切割 Alu 序列的机制，所以这些人类特异的 Alu 序列很可能反映了人类种系中新的插入，而不是黑猩猩种系中的删除。所有人类特异的 Alu 序列之间紧密的序列相似性表明它们有一个较近的共同祖先，甚至有可能存在一个“原始拷贝”Alu 序列，只有它仍然能够在人类中大量产生自身新拷贝。

人类和小鼠染色体比较表明了基因组大规模的结构趋异是如何发生的

比起人类与小鼠基因组，人类和黑猩猩基因组更加相似。虽然小鼠基因组大小几乎相同并且它含有几乎完全一样的基因，但是累积变化的时期更长（小鼠将近 10 亿年，而人类为 500 万年）。也有可能啮齿类拥有显著高于人类的突变率；在这些情况下，人类和小鼠基因组之间巨大的分歧可能以啮齿类种系中快速的序列改变占主导地位。然而，突变率种系特异的差别难以可靠地加以估计，并且在当今生物中观察到的它们对序列趋异模式所起到的作用仍然引起广泛的争论。

正如图 7-110 中 DNA 序列比较所表明的那样，突变导致人类和小鼠之间在所有非选择压力下的位点存在广泛的序列趋异性，如内含子核苷酸序列。实际上，人类-小鼠序列比较比人类-黑猩猩序列比较能够提供更多的有关基因的功能限制信息。在后者比较的情况下，几乎所有序列位置都相同，仅仅是因为最近的一个共同祖先之后没有足够的时间让大量的变化发生。相比之下，由于人类-小鼠比较中的功能限制，基因中的外显子俨然是内含子海洋中保守的小岛。

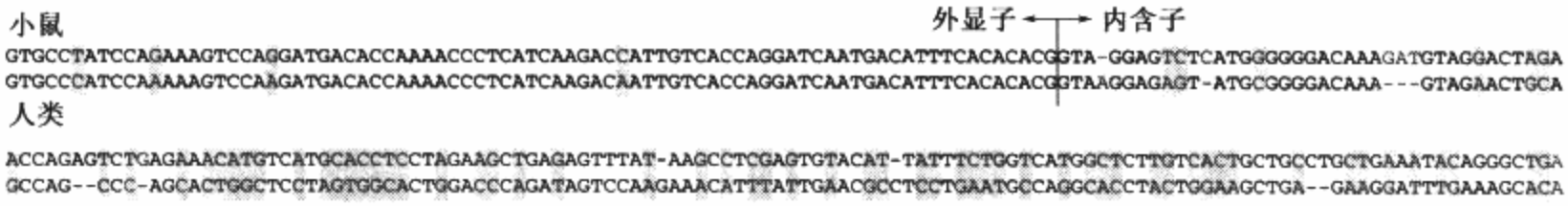


图 7-110　部分小鼠瘦蛋白和人类瘦蛋白基因的比较

绿色框表示只有一个核苷酸取代差异的序列位置，黄色表示其他的插入或删除核苷酸差异的位置。注意外显子的编码序列比相邻内含子序列要保守得多

随着完成测序的基因组不断增加，比较基因组分析正成为越来越重要的鉴定它们重要的功能位点的方法。例如，相关性很小的生物之间可读框的保守性比任何一个基因组的电脑分析都能提供更强有力的证据，来表明这些序列确实是表达基因的外显子。将来，复杂基因组（如人类和小鼠的基因组）的生物学注释将很大程度上依赖于鉴定在多个相关性较远的哺乳动物基因组中都保守的序列特征。

与人类和黑猩猩中的情况相比，局部基因序列和整体染色体组织在人类和小鼠之间已大大趋异了。根据粗略的估计，自从人类和小鼠这两个物种共有最后一个祖先以来，在这两个种系中总共发生了大约 180 次断裂和再连接事件。在此过程中，虽然这两个物种的染色体数目相似（人类单倍体基因组是 23 条染色体，而小鼠为 20 条），但它们整体结构相差很大。例如，尽管在大多数人类染色体上，中心粒占据了相对居中的位置，但它们在小鼠中靠近每个染色体的一个末端。然而，即使在大量的基因洗牌之后，仍然存在许多 DNA 大片段，其中的基因顺序在人类和小鼠中是相同的。染色体中存在这些保守基因顺序的区域称为同线性区（图 4-18）。

人类和小鼠中转座子家族的分析提供了其他有关长期趋异将这两个物种分开的证据。虽然人类主要的反转座子在小鼠中有对应物，如人类 Alu 重复在序列和转座机制方面与小鼠 B1 家族相似，但这两个家族在两个种系中独立进行扩增。即便在那些人类和小鼠序列保守性大到足以允许可靠重排的区域，人类基因组 Alu 元件的位置与小鼠基因组中 B1 元件所处对应片段还是没有相关性（图 7-111）。

难以重构远古基因组的结构

可以对远古生物的基因组进行推测，却绝不可能直接观察到：今天已没有存活的远古生物。虽然诸如马蹄蟹这种生物看起来与 20 亿年前生长的化石祖先显著相似，但完全有理由相信马蹄蟹在此期间一直和其他进化种系一样以相同的速率进行着改变。选择性限制必定维持着马蹄蟹基因组重要的功能特征，从而导致种系的形态稳定性。然而，基因组序列揭示出经历纯化选择的基因组只是一小部分；因此现代马蹄蟹的基因组必须与其灭绝的祖先有很大的差别，这些我们只能通过化石记录了解到。

早已绝种的生物基因组即便是大体上的特征也难以推测。一个重要的例子就是所谓的内

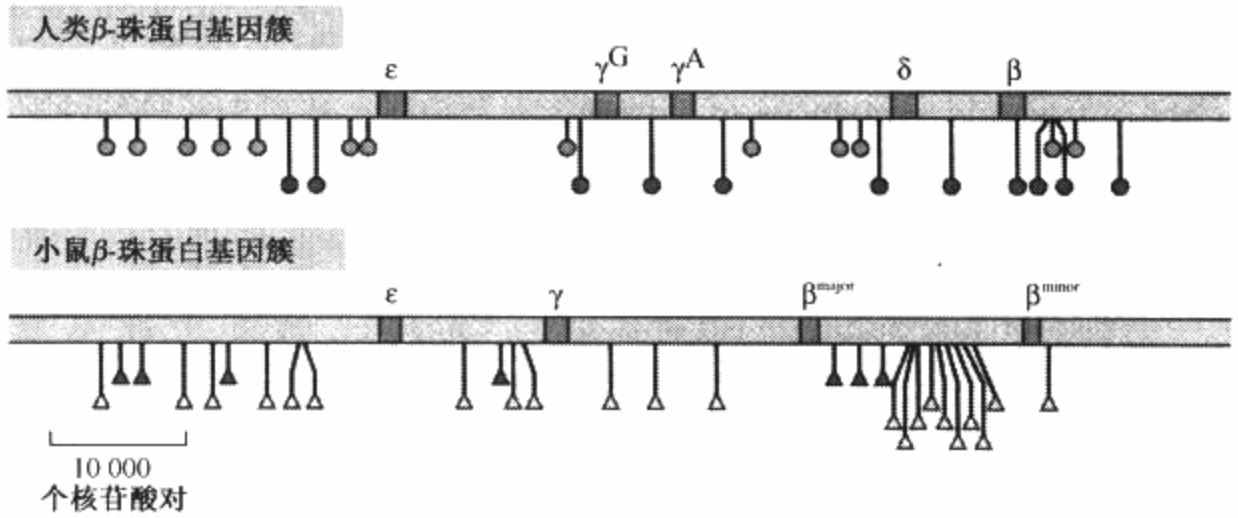


图 7-111　人类和小鼠基因组中 β-珠蛋白基因簇的比较，显示了转座子元件的位置

这段人类基因组含有 5 个功能 β-珠蛋白类似的基因（橙色），而小鼠中同类序列中只含有 4 个。绿色环表示人类 Alu 序列所在位置。红色环表示人类 L1 序列所在位置。小鼠基因组含有不同但相关的转座元件：B1 元件所在位置用蓝色三角符号表示（它与人类 Alu 序列相关），小鼠 L1 元件（与人类 L1 序列相关）所在位置用黄色三角符号表示。珠蛋白结构基因中转座元件的缺少归功于纯化选择，它可能已将任何破坏基因功能的插入删除掉（经 Ross Hardison and Webb Miller 许可）

含子早生与内含子后生的争论。1977 年发现后生动物的编码序列被内含子隔开之后不久，就出现了有关内含子是反映了地球生命进化过程中后天获得的，还是在最早的基因中就存在的争论。根据内含子早生模型，迅速生长的生物，如细菌失去存在于它们的祖先中的内含子，因为它们要选择出一个适应迅速复制的紧密基因组。这种观点受到内含子后生模型的挑战，在这种模型中内含子被看作是在单细胞生物进化很久以后，可能是通过某种类型转座子的作用，插入到无内含子的基因中。

目前还没有可靠的方式解决这个争论。现存基因组的比较研究能提供有关不同进化种系中内含子获得和丢失速率的估计。然而，这些估计只是间接与几十亿年前基因组是如何组织起来的各个问题有关。细菌和人类都是同样“现代”的生物，两者的基因组与它们最近的一个共同祖先的基因组差别如此巨大以至于我们只能推测这个非常古老的始祖基因组的特征。

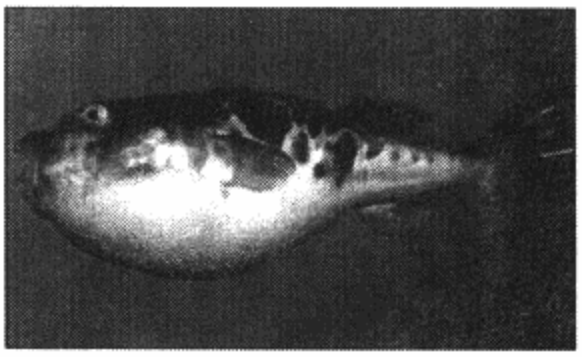


图 7-112　河豚

（经 Byrappa VenKatesh 许可）

当两种现代生物的基因中共有几乎相同的内含子位置模式的时候，我们能确定这两个物种的最近共同祖先中存在这些内含子。对人类和河豚进行比较能说明这一点（图 7-112）。河豚基因组不寻常的地方就是对于一个脊椎动物来讲它异常得小（4 亿个核苷酸对，相比之下，许多其他鱼类有 10 亿个或更多，典型哺乳动物有 30 亿个）。河豚基因组不大，几乎完全是因为其内含子很小。特别是河豚内含子以及河豚基因组其他编码片段，缺少在大多数经过详细研究的脊椎动物构成基因组大部分的重复 DNA。然而，相对于它们在哺乳动物基因组中的位置来说，河豚内含子的位置几乎完全地保守（图 7-113）。

关于河豚内含子为什么这么小的问题让人想起内含子早生与内含子后生争论。显然，或者是内含子在许多种系中生长而在河豚种系中保持很小，或者是河豚种系发生了重复序列从其内含子中大规模丢失。我们已清楚地了解了基因组如何能够通过活跃转座生长，因为大部

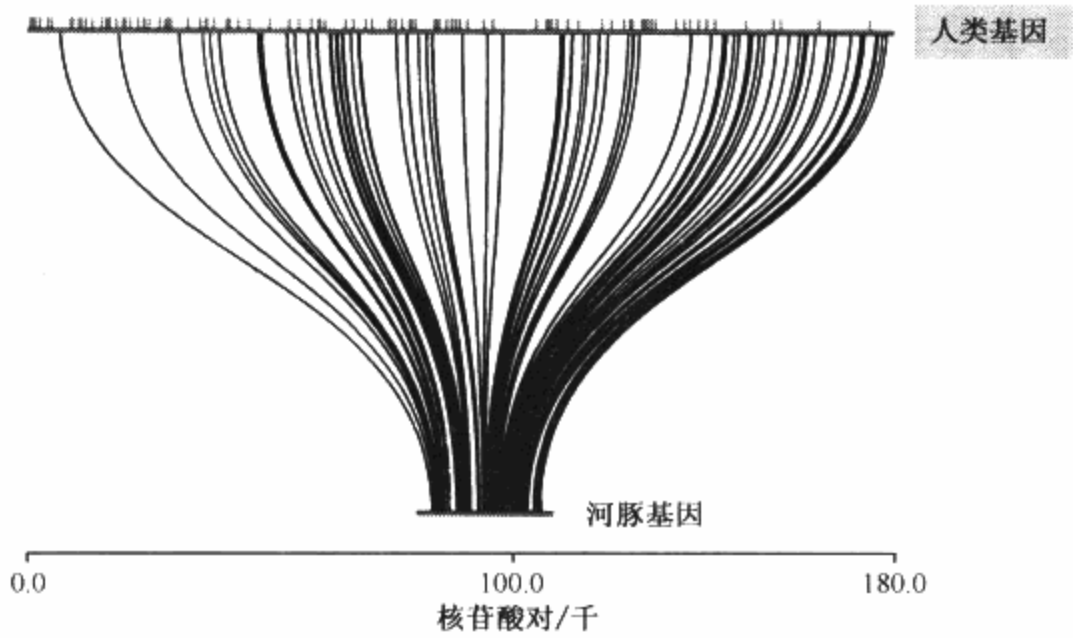


图 7-113　编码了亨廷顿蛋白的人类基因和河豚基因的基因组序列比较

两个基因（用红色表示）都含有 67 个短外显子，它们彼此按 1∶1 的比例对应排列；这些外显子用曲线连接起来。人类的基因比河豚的基因大 7.5 倍（180 000∶24 000 个核苷酸对）。大小差别全部是因为人类有更大的内含子。人类内含子较大一部分是因为存在反转录转座子，它们的位置用绿色垂直线表示；河豚内含子缺少反转录转座子。在人类中亨廷顿基因突变引起亨廷顿病，一种遗传性神经退行性紊乱（摘自 S. Baxendale et al., *Nat. Genet.* 10：67～76，1995）

分转座事件是复制性的［也就是说，原始拷贝保留在原处，而它的一个拷贝插入到新位点（图 5-72 和图 5-76）］。对于在经过仔细研究的生物中，以将会威胁种系生存的速率，从大量位点有效地删除转座子的同时却不删除相邻的重要功能序列的突变过程，其证据就极其少见了。然而，河豚异常小的内含子的起源仍然不清楚。

在进化过程中，基因重复和趋异是产生遗传新颖性的关键来源

到目前为止，我们有关基因组进化的讨论已经强调了中性变化过程或者纯化选择的影响。然而，基因组进化最重要的特征是基因组变化能够产生在进化过程中可被正向选择的生物学新颖性，从而产生了新生物类型。

比较看起来非常不同的生物后阐明了遗传新颖性的一些来源。这些比较的一个显著特征就是种系特异基因（例如，在灵长类中存在，在啮齿类中却不存在的基因，或者是在哺乳动物中发现，在其他脊椎动物中却未发现的基因）相对很少。比这要显著得多的是先存基因家族的选择性扩增。在人类、线虫、果蝇这些已完全测序的生物中编码核内激素受体的基因证明了这一点（图 7-114）。在所有这 3 种生物中，这些核内受体（也称作细胞内受体）的许多亚型有相近的同系物，它们彼此之间的相似性比它们与同一物种中存在的其他亚型之间的相似性更大。因此，整个基因大家族功能趋异有许多必须在这 3 个进化种系趋异之前进行。随后该基因家族的一个主要分支仅仅在线虫中经历了大量的扩增。特定亚型的类似但更小的种系特异性扩增在整个基因家族系统树中都是明显的，而且它们在人类中特别明显，表明这种扩增提供了通向增加生物学复杂性的途径。

基因重复看来在所有进化种系中都以高频率发生。在许多不同真核基因组中对重复基因趋异的丰度和频率进行检验后表明，任何特定基因发生成功的重排事件（也就是说一个物种大多数或全部个体都发生重复）的可能性大约是每 100 万年有 1%。对基因重复的精确机制

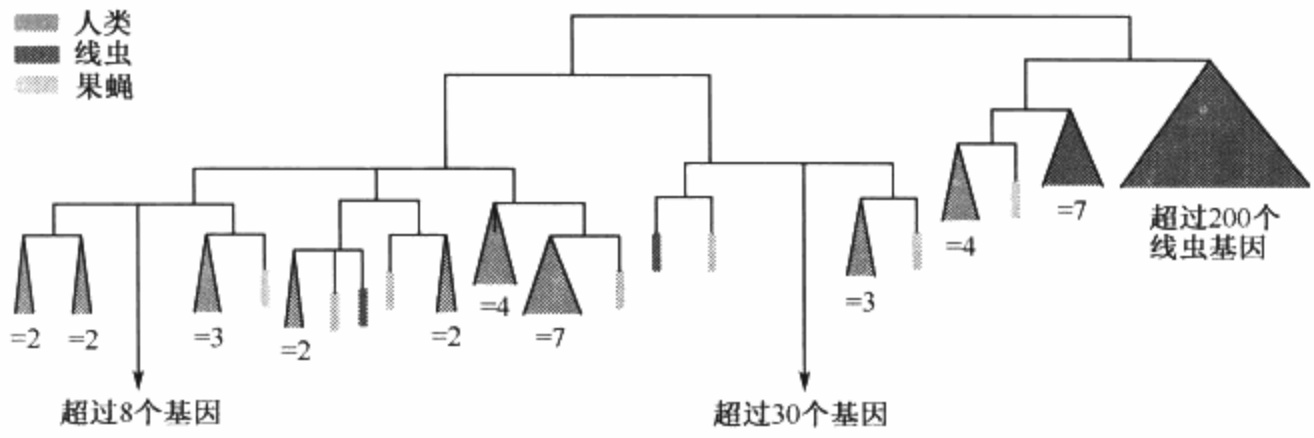


图 7-114　基于人类（*H. sapiens*）、线虫（*C. elegans*）和果蝇（*D. melangaster*）基因组中编码的所有核内激素受体推测出的蛋白质序列的系统树

三角形代表单个进化种系中发生扩增的蛋白质亚家族；这些三角符号的宽度表示编码这些亚家族成员基因的数量。彩色的垂直条带表示单个基因，对于过去发生的导致产生编码核内受体基因家族的重复和趋异事件，在 3 种现代生物中不存在它们的简单模式。图 7-14 中显示了特定核内激素受体的部分结构，第 15 章则对它们的功能进行了一般描述（摘自国际人类基因组测序共同体，*Nature* 409：860～921，2001）

所知甚少。然而，由于重复导致基因的两份拷贝常常紧挨着排列，所以有人认为双链染色体断裂的不精确修复导致重复频繁地发生（图 5-53）。

重复的基因发生趋异

基因组进化中一个主要的问题涉及新重复基因的命运。在大多数情况下，根据推测很少或者没有选择（至少最初的时候）来维持重复状态，因为任何一个拷贝都能提供相同的功能。因此，许多重复事件发生之后可能在某个或其他基因中发生功能丢失突变。这个循环将在功能上恢复重复以前的单基因状态。事实上，在目前的基因组中有很多例子，能在其中看到一个重复基因的拷贝之一因多次突变已经不可逆地失活了。长期以后，可以想到这种假基因和因重复产生该**假基因**的功能基因之间的序列相似性，会由于假基因中累积许多突变而逐渐减少，最终变得难以发现。

基因重复命运的另一选择就是在它们序列和表达模式发生趋异并发挥不同的作用时两个拷贝仍然保持其功能。这个“重复和趋异”的过程几乎肯定能解释生物学上复杂的生物中许多功能相关的基因大家族的存在，并且它被认为在更复杂生物的进化中发挥着关键的作用。

全基因组重复是重复-趋异循环中特别显著的例子。全基因组重复可以非常简单地进行：所需要的只是在生殖细胞中发生一轮基因组复制却不进行相应的细胞分裂。最初，染色体数目只是加倍了。生物**倍性**的迅速增加经常见到，尤其是在真菌和植物中。全基因组重复之后，所有基因都以重复的拷贝存在。然而，除非重复事件发生得很晚，以至于几乎没有时间在基因组结构中发生随后的变化，否则的话，会很难区分片段重复的结果（在不同的时期发生）和全基因组重复的终产物。例如，对于哺乳动物来讲，是发生了全基因组重复，还是发生 DNA 片段一系列的逐步重复就很不确定了。然而，在很早以前发生了大量基因重复却是很清楚的。

在斑马鱼中，几亿年前发生了全基因组重复或者一系列更加局部性的重复，对其基因组的分析使我们对基因重复和趋异的过程有了一些新的认识。虽然斑马鱼基因的许多重复似乎已经由于突变而丢失了，但是很大一部分（可能多达 30%～50%）在功能上已趋异了，尽管两份拷贝都保持着活性。在很多情况下，重复基因之间最明显的功能差异在于它们在不同组

织或发育的不同阶段表达（图 21-45）。解释这种结果的一种很吸引人的理论认为在一组重复基因的两份拷贝中都迅速发生着不同且轻度有害的突变。例如，由于调节型突变有一个拷贝可能在特定组织中失去表达，而另一个拷贝在另一种组织中失去表达。发生这些事件以后，就会需要两份基因的拷贝来提供以前由单个基因提供的全部功能；因此，现在两份拷贝都会通过使突变失活来防止丢失。经过更长的时期后，每份拷贝就会发生进一步的改变，由此得以获得新的特征。

珠蛋白基因家族的进化表明了 DNA 重复是怎样帮助生物进化的

珠蛋白基因家族提供了 DNA 重复如何产生新蛋白质的一个特别好的例子，因为它的进化历史已经研究得很清楚了。现今珠蛋白氨基酸序列和结构显而易见的同源性表明它们都必须起源于一个共同的祖先，即使在哺乳动物基因组中有些现在由非常不同的基因编码。

考察位于生命进化系统树上不同位置的生物中不同形式的蛋白质后，我们能够重建过去产生不同携氧血红蛋白分子类型的一些事件。由于大型动物不能依靠氧气简单地扩散通过身体表面来为它们的组织提供足够的氧气，所以就必需一种像血红蛋白这样的分子来使多细胞动物能够长成一个巨大的个体。因此，在所有脊椎动物和许多非脊椎动物中发现了血红蛋白类似分子。动物中最原始的携氧分子是一种含有约 150 个氨基酸的珠蛋白多肽链，在许多海生蠕虫、昆虫或原始鱼类中都可以发现它们的存在。然而更高等的脊椎动物中血红蛋白分子由两种珠蛋白链组成。大约 50 亿年前，在高等鱼类进化过程中发生了一系列基因突变和重复。这些事件产生了两种略为不同的珠蛋白基因，在每个个体的基因组中分别编码 α 和 β 珠蛋白链。在现代的高等脊椎动物中，每种血红蛋白分子是由两条 α 链和两条 β 链组成的复合体（图 7-115），这个 $\alpha_2\beta_2$ 分子中 4 个氧结合位点相互作用，使该分子在结合以及释放氧气的时候分子内能发生协同性异构改变，从而使血红蛋白能比单链的形式更有效地携带和释放氧气。

再后来，哺乳动物进化的过程中，β 链基因显然发生了重复和突变，从而又产生一种 β 类似链，且特异地在胎儿中合成。结果生成的血红蛋白分子对氧的亲和力比成年动物的血红蛋白更高，因而帮助从母亲输送氧气给胎儿。编码新的 β 类似链的基因随后再次突变和重复，以产生两个新的基因，ε 和 γ，其中 ε 链在发育过程（将形成 $\alpha_2\epsilon_2$）中比胎儿 γ 链合成更早，后者将形成 $\alpha_2\gamma_2$。在灵长类动物的进化过程中，成年 β 链基因发生重复更晚一些，以产生一个 δ-珠蛋

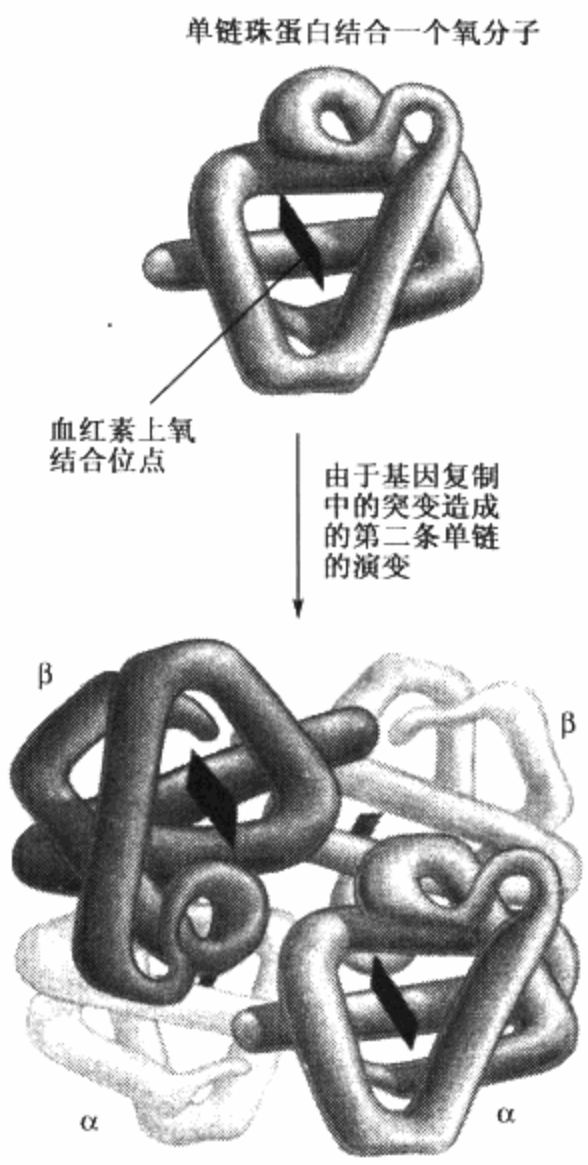


图 7-115　单链和 4 条链的珠蛋白结构比较
所显示的 4 条链的珠蛋白是血红蛋白，它是由两条 α 链和两条 β 链形成的复合体。在有些原始脊椎动物中一条链的珠蛋白形成一个二聚体，当它结合氧气后就发生解离，代表了 4 条链珠蛋白进化过程中的中间物

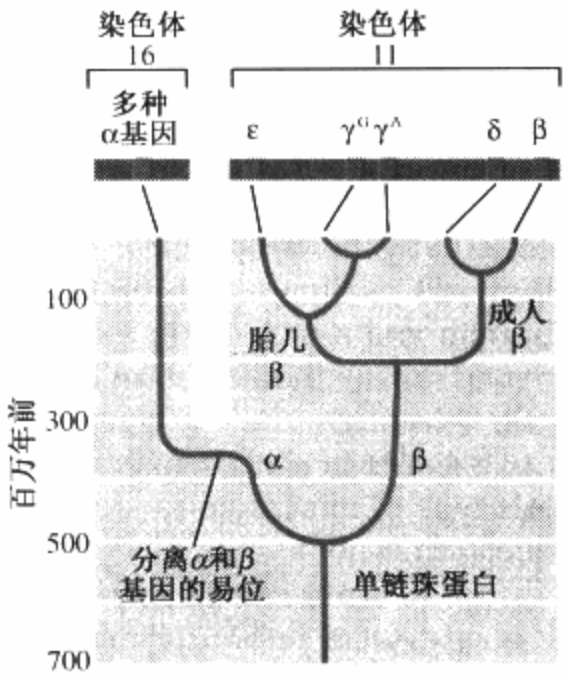


图 7-116　在动物血液中运输氧气的珠蛋白链进化体系

该体系强调了 β 类似珠蛋白基因家族。γ 链基因相对较晚发生的基因重复产生了 γ^G 和 γ^A，它们是具有相同功能的胎儿 β 类似链。图的顶部显示的是珠蛋白基因在人类基因组中的位置（图 7-60）

白的基因并因此产生一种较小的只在成年灵长类动物中发现的血红蛋白形式（$\alpha_2\delta_2$）（图 7-116）。

这些重复基因每个都被影响最终血红蛋白分子功能的点突变，以及决定该基因表达时间和水平的调控区中发生的改变所修饰。结果，每个珠蛋白在人类发育的不同时期都以不同的量合成[图7-60(B)]。

在由起初的 β 基因衍生的人类基因中，可以清楚地看到导致产生不同珠蛋白链的基因重复过程的最终结果，这些基因作为一系列同源 DNA 序列彼此排列在 50 000个核苷酸对中。类似的一簇 α-珠蛋白基因位于另外一条人类染色体上。由于在鸟类和哺乳动物中 α-和 β-珠蛋白基因簇位于不同的染色体上，但在非洲爪蟾中位于一条染色体上，所以人们相信大约 30 亿年前染色体转位事件将两个基因簇分开了（图 7-116）。

在不是功能基因而是假基因的 α-和 β-珠蛋白基因簇中存在多个重复的珠蛋白 DNA 序列。它们与功能基因有很近的同源性，但是因突变阻止它们的表达而失活了。正如所期望的那样，这种假基因的存在清楚地表明不是每种 DNA 重复都产生新的功能基因。我们还知道非功能 DNA 序列并没有被迅速丢弃，哺乳动物基因组中发现大量冗余的非编码 DNA 可以表明这一点。

外显子重组可以产生编码新蛋白质的基因

DNA 重复在进化中的作用并不限于扩增基因家族。它也能小规模地作用，通过将短 DNA 重复片段串联起来产生单独的基因。以这种方式产生的基因所编码的蛋白质可以通过重复且类似的蛋白质结构域的存在加以识别，它们彼此共价串联起来。例如，免疫球蛋白（图 7-117）和白蛋白，以及大多数纤维状蛋白（如胶原蛋白）都是由通过原始 DNA 序列的反复重复进化得到的基因编码。

以这种方式进化的基因，以及许多其他基因中，每个单独的外显子编码单个蛋白质折叠单位或结构域。有人相信 DNA 编码序列以这种长内含子将一系列外显子分隔开的形式组织起来，大大促进了新蛋白质的进化。例如，形成一个具有重复结构域的蛋白质编码基因所必需的重复，可以通过将编码有用蛋白质结构域的外显子任一侧的长内含子中任何地方的 DNA 断裂和再连接来实现；如果没有内含子，原来基因中就只有几个 DNA 分子间的重组交换能重复该结构域的位点。通过使重复能够由重组在许多潜在位点而不是几个位点上进行，内含子增加了有利重复事件。

从更一般的意义上讲，我们从基因组序列认识到，基因的组成部分，无论是它们的单个外显子还是它们的调控元件，都作为模块单元，在基因组内重复和四处移动，从而形成了当今生命丰富的多样性。结果，许多现今蛋白质由来自不同结构域家族的结构域拼凑而成，反映了它们长久的进化历史（图 7-118）。

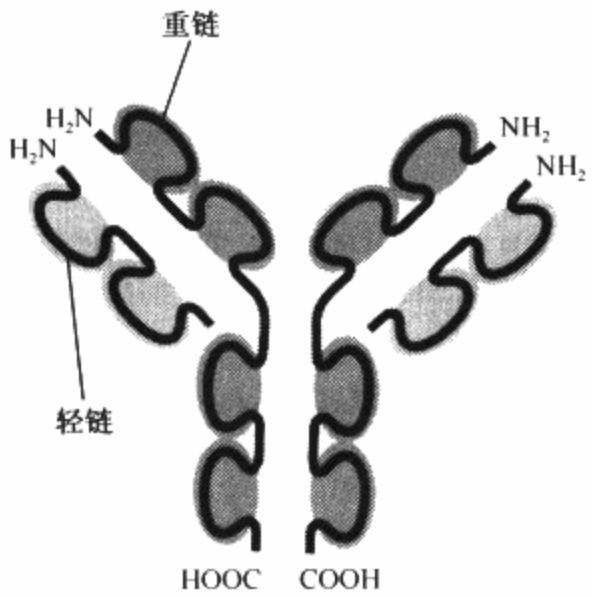


图 7-117　抗体分子（免疫球蛋白）简图

该分子是由两条相同的重链和两条相同的轻链形成的复合体。每条重链含有 4 个相似的共价结合的结构域，而每条轻链含有两个这种结构域。每个结构域由单独的外显子编码，所有的外显子据认为由同一外显子始祖经过一系列重复后进化而来

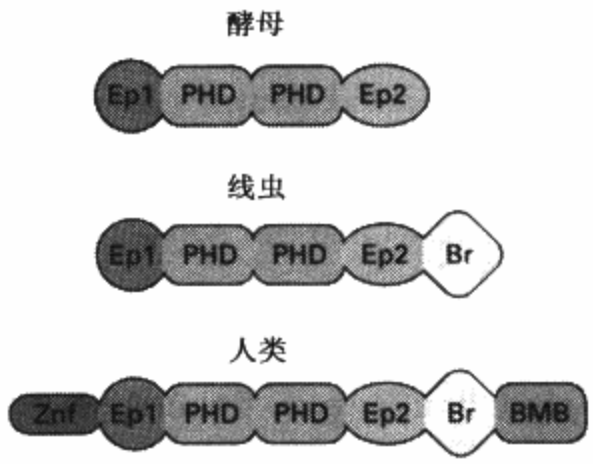


图 7-118　一组据认为有类似功能的进化相关蛋白质的结构比较

一般来讲存在一种趋势，就是越复杂生物中的蛋白质，如我们自己，越可能含有额外的结构域，这里比较的 DNA 结合蛋白也是如此

基因组序列给科学家留下尚待解开的谜

既然我们从基因组序列知道了人和小鼠拥有本质上相同的基因，我们就不得不面对在 21 世纪对细胞生物学家提出挑战的一个主要问题。考虑到人和小鼠由相同的一套蛋白质形成，那么在进化过程中发生了什么事件，导致小鼠和人类如此不同呢？虽然目前这个问题的答案存在于每个完成测序的基因组 30 亿个核苷酸序列中每个地方，但我们还不知道如何去破译这种信息，所以这个关键而又非常基本的问题的答案还不清楚。

尽管有很多我们还不知道，但是可能值得去做一些推测，只要能有助于指明通向解决前面难题的道路。在生物学中，时间的选择非常重要，当我们检验使受精卵发育成胚胎，胚胎再发育成成年个体的复杂机制时，这一点就会很清楚（讨论见第 21 章）。在我们的发育过程中，有关哪一个 RNA 分子和哪一个蛋白质将被合成，它们各自在什么地方，以及到底什么时候和以多大量合成的几十亿个决定的结果是形成了人体。比起黑猩猩或小鼠，这些决定对人类来说是不同的。基因组编码序列代表一套或多或少是标准的 30 000 个左右的基本组分，由它们形成这 3 种生物。因此在很大程度上必定是本章描述的基因表达的许多不同调控方式产生了人类和其他哺乳动物之间的差异。

考虑到这些假设，期望基因组以某种允许生物在经过选择的细胞中尝试改变的基因定时和表达模式的方式得以进化将会更加合情合理。当我们讨论选择性 RNA 剪接和 RNA 编辑机制时，我们已经看到一些证据表明确实如此。似乎也存在一些机制（某些基于转座 DNA 元件移动的机制）允许模块轻易地加入到基因调控区或者从基因调控区删除，以便在生物进化时改变它们的转录模式。实际上，对这些调控区的分析提供证据支持这样的观点，即认为大部分基因调控区由基因调节蛋白识别的 DNA 结合位点在进化中混合和搭配而成（图 7-119）。

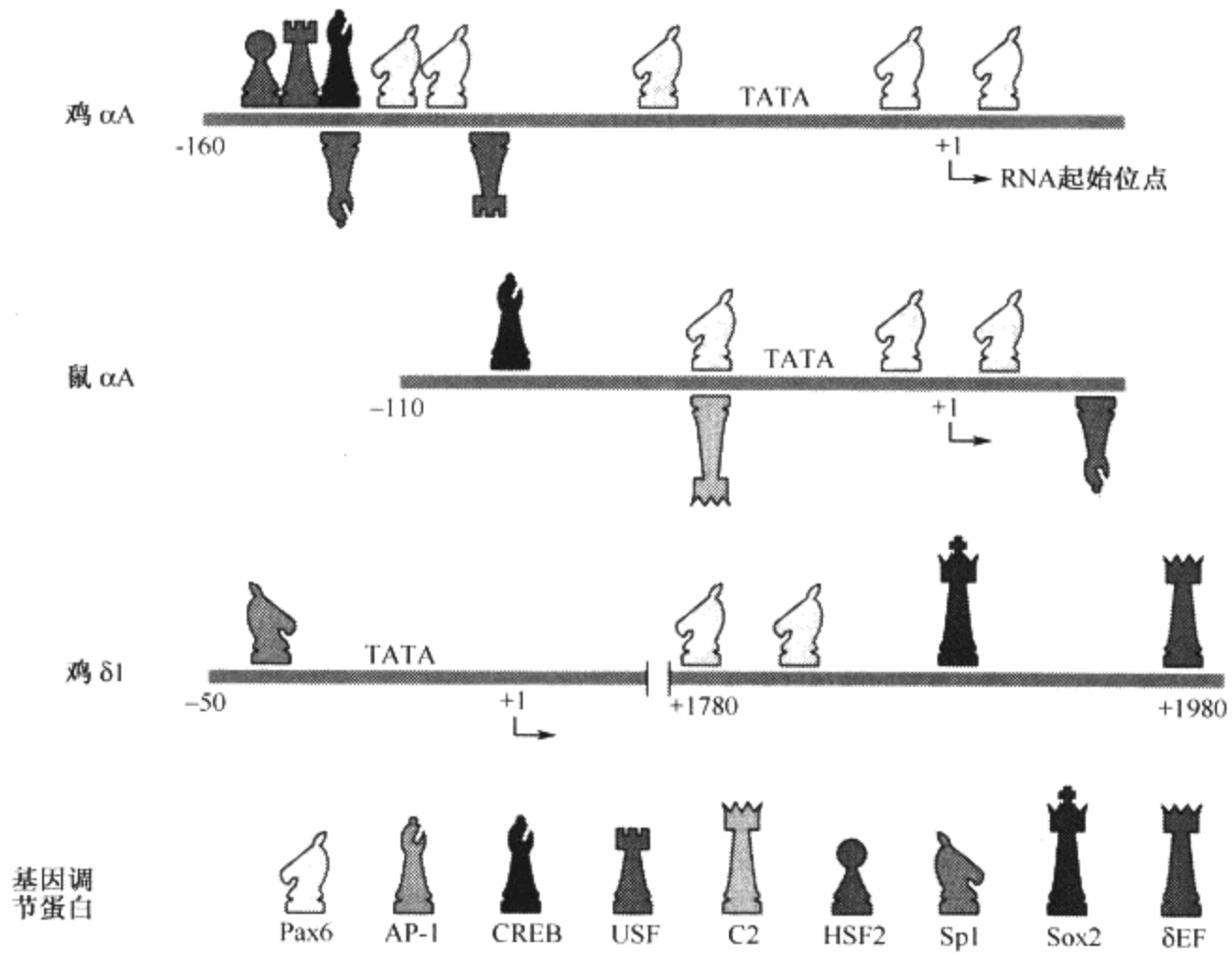


图 7-119　小鼠和雏鸡眼部晶状体的基因调控区

晶体蛋白组成了大量的晶状体并负责将光折射或聚焦到视网膜上。细胞内的许多蛋白质有特性（高溶解度、合适的折射系数等），适合晶状体的功能，并且在进化过程中选择了多种这类蛋白质用做晶状体。例如，α 晶体蛋白（上面的两条直线）与热激蛋白密切相关，并在所有脊椎动物晶状体中都能找到。相比之下，δ 晶体蛋白（第三条直线）与一个参与氨基酸代谢的酶密切相关，且只在鸟类和两栖类动物中发现。图中所示的 3 种晶体蛋白基因调控区是由反映了各个基因进化历史的不同调节序列组成的拼装物。所有 3 种调控区的共同特征是存在基因调节蛋白 Pax6 的结合位点。Pax6 是果蝇 Toy 和 Eyless 蛋白的脊椎动物同系物（图 7-75），也是指令眼部发育的重要调节蛋白。每个基因调控区上面的蛋白质都是转录激活蛋白，直线下面的则是阻遏蛋白（摘自 E. H. Davidson, Genomic Regulatory system: Development and Evolution, pp. 191～201. San Diego: Academic press, 2001 and A. Cvekl and J. Piatigarsky, BioEssays 18: 621～630, 1996）

物种的遗传变异为基因组的进化提供了一幅精细的画面

对比两个已彼此趋异达几百万年的物种时，用每个物种的哪一个个体来做比较并没有什么分别。例如，典型的人类和黑猩猩 DNA 序列彼此相差 1%。相比之下，如果从两个不同的人抽取基因组相同区域进行比较，差别通常不到 0.1%。对于相关性更远的生物而言，种间差异甚至更加显著地掩盖了种内变化。然而，人类和黑猩猩之间每个固定的差异（也就是现在已成每个物种所有或几乎所有个体特有的差异）开始成为单个个体中新的突变。如果发生突变的种间群体大小是 N，那么对于二倍体生物，新突变原先的等位基因频率将是 $1/2N$。这种稀有突变是怎样在群体中固定，并因而成为该物种的特征而不是特定个体基因组特征的呢？

这个问题的答案取决于突变的功能结果。如果突变产生非常有害的影响，它将只是被纯化选择删除，并且将不会固定下来（在最极端的情况下，携带突变的个体将死亡且不会产生后代）。相反，赋予继承它们的个体大部分繁殖优势的稀有突变将迅速在群体中散布开来。因

为人类是有性生殖的，并且每次形成合子时都会发生遗传信息重组，所以每个继承突变的个体基因组都是由继承自大量祖先的片段组成的独一无二的重组镶嵌物。经过选择的突变以及适度数量的相邻序列，最终从发生突变的个体中继承得到的，将只是这个巨大镶嵌物的一小块而已。

大多数无害的突变也不会有用。这些选择性的中性突变也能散布，并在群体中固定下来，它们为基因组中的进化变异做出了大量贡献。它们的散布速度并不像稀有且十分有利的突变那样快。这种中性遗传变异在理想化的杂交繁殖群体遗传的过程可以用令人惊奇的简单数学方程表示。这个经过证明对于分析人类遗传变异非常有用的理想模型对突变做出几个假设，包括群体大小恒定、随机交配以及选择性中性。尽管这些假设里面没有一个能很好地描述了人类种群历史，然而它们为分析种内变化提供了一个有用的起点。

在一个进行着随机交配，大小为 N 的恒定群体中，如果发生了新的中性突变，那么它将被固定下来的可能性大约是 $1/2N$。对于那些已经被固定的突变而言，固定所用的平均时间大约是 $4N$ 个世代。对人类遗传变异数据详细分析后表明，在大部分目前的遗传变异模式建立期间，远古群体大小约为 10 000。在这些情况下，一个新的选择性中性突变被固定下来的可能性很小（5×10^{-5}），尽管平均固定时间在 800 000 年。因此，虽然我们知道自从大约15 000 年前农业发展以来，人类群体就大大增加了，但是人类群体中大多数人类遗传变异的产生和建立是在此之前发生的，那时人类群体还很小。

即使现代人类的大部分变异起源于比较少的先祖群体中存在的变异，但是所遇到的变异数目却很大。大部分变异采用单核苷酸多态性（SNP）的形式。它们仅仅是基因组序列上的位点，在这些位点上，很大一部分人群中是一种核苷酸，而另一部分是另一种核苷酸。从现代人类群体中随机抽取两个人的基因组，它们大约在 2.5×10^{6} 个位点（每 1300 核苷酸对有一个）存在差异。在人类基因组中存在多态性的作图位点（意思是在此位点两个个体基因组可能不相同是合理的）对于遗传分析极其有用，其中力图将特定的特征（表型）和特定 DNA 序列关联起来，以便达到医学或者科学上的目的。

某些具有极高突变率的序列与继承自史前先祖的普通 SNP 形成鲜明对比。一个显著的例子就是 CA 重复，它们在人类基因组和其他真核生物基因组中广泛分布。具有 $(CA)_n$ 模式的序列以相对较低的忠实性进行复制，因为 DNA 复制期间在模板和新合成链之间发生滑动；因此，不同基因组之间 n 的精确值可以在很大范围内变动。这些重复产生了理想的基于 DNA 的遗传标记，因为大多数人是杂合子——每个特定的 CA 重复携带两个 n 值，其中一个重复长度继承自母亲，另一个不同重复长度继承自父亲。尽管 n 值变化如此罕见以至于大多数父母-后代传递过程忠实地扩增了 CA 重复，但是这些变化频繁程度仍然足以维持人类群体中高水平的杂合状态。这些以及其他显示极高可变性的简单重复在犯罪调查、父母鉴定和其他法庭应用中提供了利用 DNA 分析鉴定个体的基础（图 8-41）。

尽管认为大多数 SNP 和人类基因组序列中其他常见变化对表型没有影响，但是它们的一部分必须负责人类个体几乎所有的可遗传方面。人类遗传学的一个主要挑战就是学会识别那些相对稀少的但在功能上又很重要的变异，以区别任何两个人基因组的中性变异为大背景。

小结

现今基因组核苷酸序列比较完全改变了我们对基因和基因组进化的认识。由于 DNA 复制和 DNA 修复过程的极高忠实性，在维持基因组核苷酸序列中很少出现随机错误，每 100 万年内每 1000 个核苷酸大约会有 5 个发生改变。因此，比较人类和黑猩猩的染色体（它们经

过500万年时间的进化后分开）后发现基本没有变化就不足为奇了。不仅我们的基因本质上是一样的，而且它们在染色体上的排列顺序也几乎相同。此外，组成我们非编码DNA大部分的转座元件的位置也大多不变。

如果将两个相距甚远的生物进行比较（如人和小鼠，经过10亿年进化后分开）就会发现更多的变化。现在可以清楚地看到自然选择的作用：通过纯化选择，重要的核苷酸序列［位于调节区和编码序列中（外显子序列）］都高度保守。相比之下，非关键序列（如内含子序列）以某种程度发生改变，以至于常常不能根据先祖进行精确的序列比对。

由于纯化选择的存在，可以跨越很远的种系发生距离识别同源基因，并且常常能构建特定基因的详细进化历史，从而根据其历史找出现今物种的共同祖先。因此我们可以看到现今生物的大量遗传复杂性是由于古老基因家族的扩增所造成的。发生趋异之前进行的DNA重复因而成为进化过程中产生遗传新颖性的主要来源。

参考文献

一般文献

Carey M & Smale ST (2000) Transcriptional Regulation in Eukaryotes: Concepts, Strategies and Techniques. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.

Hartwell L, Hood L, Goldberg ML et al. (2000) Genetics: from Genes to Genomes. Boston: McGraw Hill.

Lewin B (2000) Genes VII. Oxford: Oxford University Press.

Lodish H, Berk A, Zipursky SL et al. (2000) Molecular Cell Biology, 4th edn. New York: WH Freeman.

McKnight SL & Yamamoto KR (eds) (1992) Transcriptional Regulation. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.

Mechanisms of Transcription. (1998) *Cold Spring Harb. Symp. Quant. Biol.* 63.

An Overview of Gene Control

Campbell KH, McWhir J, Ritchie WA & Wilmut I (1996) Sheep cloned by nuclear transfer from a cultured cell line. *Nature* 380, 64–66.

Gurdon JB (1992) The generation of diversity and pattern in animal development. *Cell* 68, 185–199.

Ross DT, Scherf U, Eisen MB et al. (2000) Systematic variation in gene expression patterns in human cancer cell lines. *Nat. Genet.* 24, 227–235.

DNA-binding Motifs in Gene Regulatory Proteins

Bulger M & Groudine M (199) Looping versus linking: toward a model for long-distance gene activation. *Genes Dev.* 13, 2465–2477.

Choo Y & Klug A (1997) Physical basis of a protein–DNA recognition code. *Curr. Opin. Struct. Biol.* 7, 117–125.

Gehring WJ, Affolter M & Burglin T (1994) Homeodomain proteins. *Annu. Rev. Biochem.* 63, 487–526.

Harrison SC (1991) A structural taxonomy of DNA-binding domains. *Nature* 353, 715–719.

Jacob F & Monod J (1961) Genetic regulatory mechanisms in the synthesis of proteins. *J. Mol. Biol.* 3, 318–356.

Laity JH, Lee BM & Wright PE (2001) Zinc finger proteins: new insights into structural and functional diversity. *Curr. Opin. Struct. Biol.* 11, 39–46.

Lamb P & McKnight SL (1991) Diversity and specificity in transcriptional regulation: the benefits of heterotypic dimerization. *Trends Biochem. Sci.* 16, 417–422.

McKnight SL (1991) Molecular zippers in gene regulation. *Sci. Am.* 264, 54–64.

Muller CW (2001) Transcription factors: global and detailed views. *Curr. Opin. Struct. Biol.* 11, 26–32.

Orlando V (2000) Mapping chromosomal proteins *in vivo* by formaldehyde-crosslinked-chromatin. *Trends Biochem. Sci.* 25, 99–104.

Pabo CO & Sauer RT (1992) Transcription factors: structural families and principles of DNA recognition. *Annu. Rev. Biochem.* 61, 1053–1095.

Ptashne M (1992) A Genetic Switch, 2nd edn. Cambridge, MA: Cell Press and Blackwell Press.

Rhodes D, Schwabe JW, Chapman L et al. (1996) Towards an understanding of protein–DNA recognition. *Proc. R. Soc. Lond.* B 351, 501–509.

Seeman NC, Rosenberg JM & Rich A (1976) Sequence-specific recognition of double helical nucleic acids by proteins. *Proc. Natl Acad. Sci. USA* 73, 804–808.

Wolberger C (1996) Homeodomain interactions. *Curr. Opin. Struct. Biol.* 6, 62–68.

How Genetic Switches Work

Beckwith J (1987) The operon: an historical account. In *Escherichia coli* and *Salmonella typhimurium*: Cellular and Molecular Biology (Neidhart FC, Ingraham JL, Low KB et al. eds), vol 2, pp 1439–1443. Washington, DC: ASM Press.

Bell AC, West AG & Felsenfeld G (2001) Insulators and boundaries: versatile regulatory elements in the eukaryotic. *Science* 291, 447–450.

Buratowski S (2000) Snapshots of RNA polymerase II transcription initiation. *Curr. Opin. Cell Biol.* 12, 320–325.

Carey M (1998) The enhanceosome and transcriptional synergy. *Cell* 92, 5–8.

Fraser P & Grosveld F (1998) Locus control regions, chromatin activation and transcription. *Curr. Opin. Cell Biol.* 10, 361–365.

Kadonaga JT (1998) Eukaryotic transcription: an interlaced network of transcription factors and chromatin-modifying machines. *Cell* 92, 307–313.

Kercher MA, Lu P & Lewis M (1997) Lac repressor-operator complex. *Curr. Opin. Struct. Biol.* 7, 76–85.

Kornberg RD (1999) Eukaryotic transcriptional control. *Trends Cell Biol.* 9, M46–49.

Malik S & Roeder RG (2000) Transcriptional regulation through Mediator-like coactivators in yeast and metazoan cells. *Trends Biochem. Sci.* 25, 277–283.

Merika M & Thanos D (2001) Enhanceosomes. *Curr. Opin. Genet. Dev.* 11, 205–208.

Myers LC & Kornberg RD (2000) Mediator of transcriptional regulation. *Annu. Rev. Biochem.* 69, 729–749.

Ptashne M & Gann A (1998) Imposing specificity by localization: mechanism and evolvability. *Curr. Biol.* 8, R812–R822.

Schleif R (1992) DNA looping. *Annu. Rev. Biochem.* 61, 199–223.

St Johnston D & Nusslein-Volhard C (1992) The origin of pattern and polarity in the *Drosophila* embryo. *Cell* 68, 201–219.

Struhl K (1998) Histone acetylation and transcriptional regulatory mechanisms. *Genes Dev.* 12, 599–606.

The Molecular Genetic Mechanisms that Create Specialized Cell Types

Bird AP & Wolffe AP (1999) Methylation-induced repression—belts, braces, and chromatin. *Cell* 99, 451–454.

Cross SH & Bird AP (1995) CpG islands and genes. *Curr. Opin. Genet. Dev.* 5, 309–314.

Dasen J & Rosenfeld M (1999) Combinatorial codes in signaling and synergy: lessons from pituitary development. *Curr. Opin. Genet. Dev.* 9, 566–574.

Gehring WJ & Ikeo K (1999) Pax 6: mastering eye morphogenesis and eye

evolution. *Trends Genet.* 15, 371–377.

Haber JE (1998) Mating-type gene switching in *Saccharomyces cerevisiae*. *Annu. Rev. Genet.* 32, 561–599.

Marin I, Siegal ML & Baker BS (2000) The evolution of dosage-compensation mechanisms. *Bioessays* 22, 1106–1114.

Meyer BJ (2000) Sex in the worm: counting and compensating X-chromosome dose. *Trends Genet.* 16, 247–253.

Robertson BD & Meyer TF (1992) Genetic variation in pathogenic bacteria. *Trends Genet.* 8, 422–427.

Surani MA (1998) Imprinting and the initiation of gene silencing in the germ line. *Cell* 93, 309–312.

Weintraub H (1993) The MyoD family and myogenesis: redundancy, networks, and thresholds. *Cell* 75, 1241–1244.

Wolberger C (1999) Multiprotein-DNA complexes in transcriptional regulation. *Annu. Rev. Biophys. Biomol. Struct.* 28, 29–56.

Young MW (1998) The molecular control of circadian behavioral rhythms and their entrainment in Drosophila. *Annu. Rev. Biochem.* 67, 135–152.

Posttranscriptional Controls

Baker BS (1989) Sex in flies: the splice of life. *Nature* 340, 521–524.

Benne R (1996) RNA editing: how a message is changed. *Curr. Opin. Genet. Dev.* 6, 221–231.

Cline TW & Meyer BJ (1996) Vive la difference: males vs females in flies vs. worms. *Annu. Rev. Genet.* 30, 637–702.

Dever TE (1999) Translation initiation: adept at adapting. *Trends Biochem. Sci.* 24, 398–403.

Frankel AD & Young JAT (1998) HIV-1: fifteen proteins and an RNA. *Annu. Rev. Biochem.* 67, 1–25.

Graveley BR (2001) Alternative splicing: increasing diversity in the proteomic world. *Trends Genet.* 17, 100–107.

Gray NK & Wickens M (1998) Control of translation initiation in animals. *Annu. Rev. Cell Dev. Biol.* 14, 399–458.

Hentze MW & Kulozik AE (1999) A perfect message: RNA surveillance and nonsense-mediated decay. *Cell* 96, 307–310.

Hinnebusch AG (1997) Translational regulation of yeast GCN4. A window on factors that control initiator-tRNA binding to the ribosome. *J. Biol. Chem.* 272, 21661–21664.

Holcik M, Sonenberg N & Korneluk RG (2000) Internal ribosome initiation of translation and the control of cell death. *Trends Genet.* 16, 469–473.

Jansen RP (2001) mRNA localization: message on the move. *Nat Rev Mol Cell Biol* 2, 247–256.

Pollard VW & Malim MH (1998) The HIV-1 Rev protein. *Annu. Rev. Microbiol.* 52, 491–532.

Sharp PA (2001) RNA interference – 2001. *Genes Dev.* 15, 485–490.

Wilusz CJ, Wormington M & Peltz SW (2001) The cap-to-tail guide to mRNA turnover. *Nat. Rev. Mol. Cell Biol.* 2, 237–246.

How Genomes Evolve

Dehal P, Predki P, Olsen AS et al. (2001) Human chromosome 19 and related regions in mouse: conservative and lineage-specific evolution. *Science* 293, 104–111.

Henikoff S, Greene EA, Pietrokovski S et al. (1997) Gene families: the taxonomy of protein paralogs and chimeras. *Science* 278, 609–614.

International Human Genome Sequencing Consortium (2001) Initial sequencing and analysis of the human genome. *Nature* 409, 860–921.

Kumar S & Hedges SB (1998) A molecular timescale for vertebrate evolution. *Nature* 392, 917–920.

Li WH, Gu Z, Wang H & Nekrutenko A (2001) Evolutionary analyses of the human genome. *Nature* 409, 847–849.

Long M, de Souza SJ & Gilbert W (1995) Evolution of the intron-exon structure of eukaryotic genes. *Curr. Opin. Genet. Dev.* 5, 774–778.

Rowold DJ & Herrera RJ (2000) Alu elements and the human genome. *Genetica* 108, 57–72.

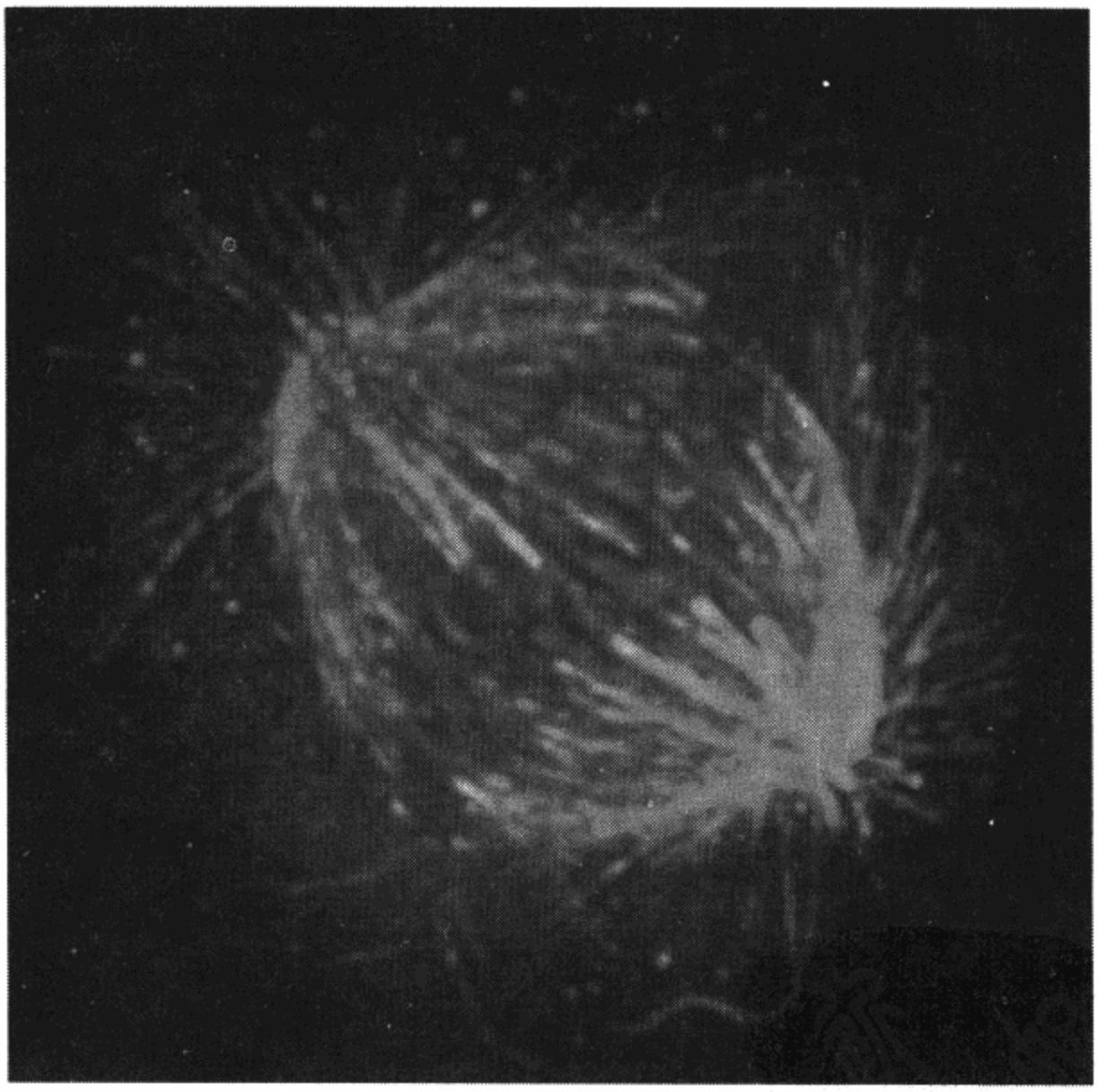
Stoneking M (2001) Single nucleotide polymorphisms. From the evolutionary past. *Nature* 409, 821–822.

Wolfe KH (2001) Yesterday's polyploids and the mystery of diploidization. *Nat. Rev. Genet.* 2, 333–341.

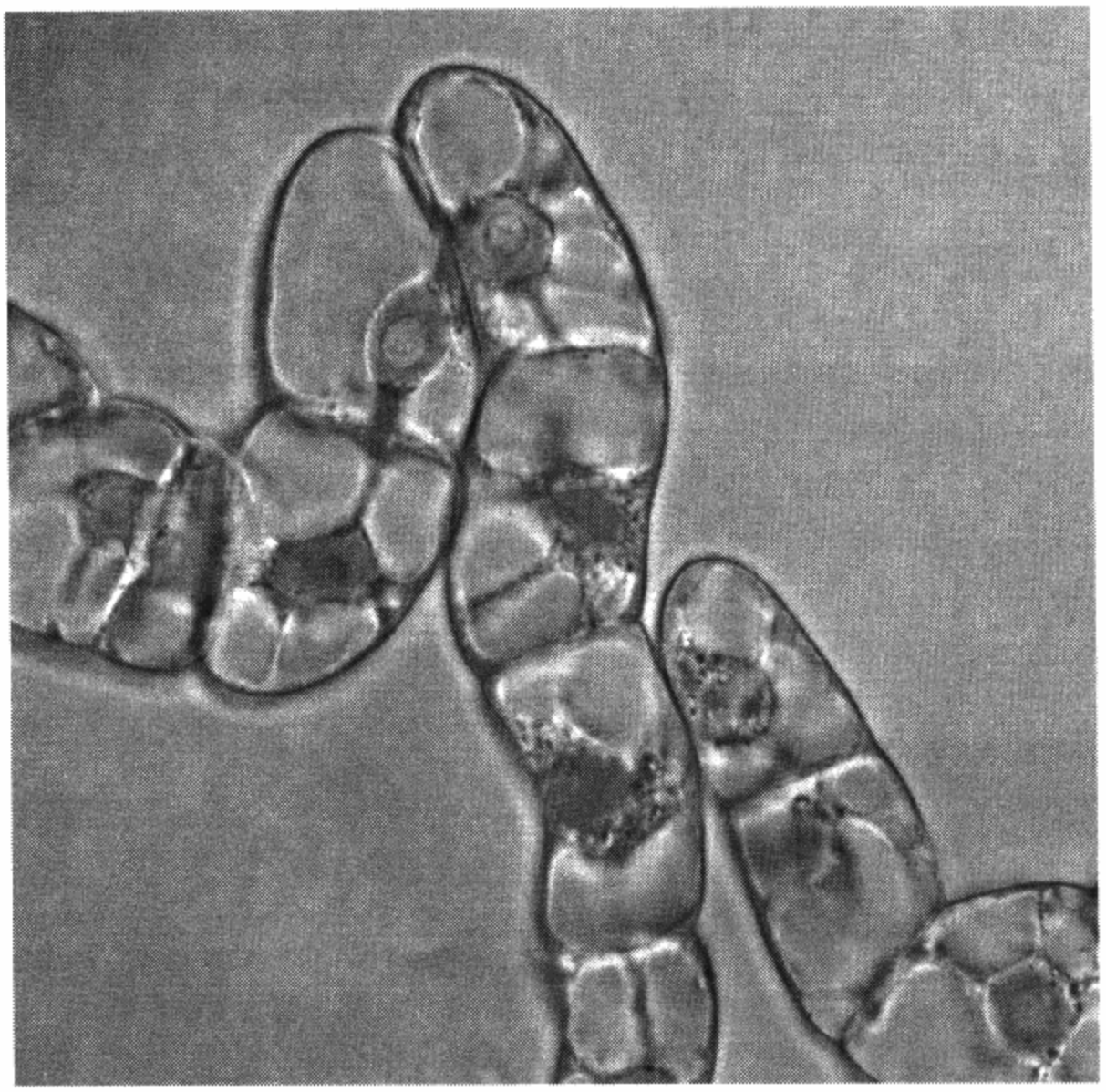
第 III 部分　方　　法

8　操纵蛋白质、DNA 和 RNA

9　细胞显像



荧光显微照相。新的探针和改进的荧光显微镜使我们辨别不同细胞组分和精确鉴定它们亚细胞定位的能力发生了变革。为了展示出纺锤体的微管，图中将有丝分裂纺锤体染成了绿色，将染色体中的 DNA 染成了蓝色，将着丝粒染成了红色（经 Kevin F. Sullivan 许可）



培养基中的植物细胞。这些烟草细胞生长在液体培养基中。一个标记有绿色荧光蛋白质的组蛋白被掺入染色质中。其中一些正在分裂，呈现出压缩了的染色体。不论是动物细胞还是植物细胞的培养基都是细胞生物学重要的研究工具（经 Gethin Roberts 许可）

8　操纵蛋白质、DNA 和 RNA

细胞分离和培养
细胞组分的分离
分离、克隆 DNA 和 DNA 测序
分析蛋白质的结构和功能
研究基因的表达和功能

技术的进步常常驱策科学的进步。例如，在生物学领域里，荷兰商人安顿·凡·列文虎克（Anton van Leeuwenhoek）创造了最早的显微镜镜头。正是通过不可思议的新镜片，列文虎克发现了一个在此之前从未被看到过的细胞的世界。在这个世界中，微小的生物体能在一滴水中翻转旋转（图 8-1）。

21 世纪注定会成为生物学异常活跃的时代。分析蛋白质、DNA 和 RNA 的新方法促成信息的爆炸，使科学家们能够以从前难以想像的方式去研究细胞和其中的大分子。我们现在能够得到数以 10 亿计的核苷酸序列，它们承载着数十种生命体的完整的分子蓝图，从微生物和芥菜到蠕虫，从果蝇到人类。强有力的新技术还帮助我们破译这些信息，不仅是我们得以编辑有关基因和蛋白质的庞杂的、详尽的数据，而且帮助我们开始搞清这些组分是如何共同组成有功能的细胞和生物体的。我们的目标可以归纳为完全弄清一个细胞在对环境做出反应和与相邻细胞发生相互作用时细胞内所发生的一切。我们想知道哪个基因被启动，哪种 mRNA 发生了转录，什么蛋白质产生了活性，它们在什么位置，它们与谁共同作用，它们属于哪条信号途径或信号网。我们还想知道细胞是如何成功地处理其惊人的数量和可变性的，如何从几乎无穷的可能性中选择的生物学功能。对这些信息的获得将使我们能够开始建立一个知识体系，并最终预言基因和蛋白质如何构成生命的基础。

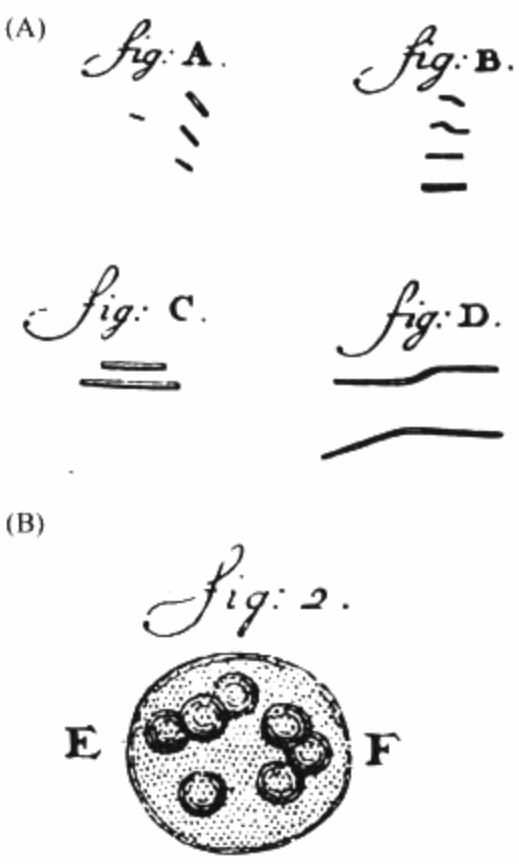


图 8-1　显微镜下的生命
列文虎克用他简陋的显微镜做的“多样的微生物”的示范。（A）从他的牙缝中刮下来的细菌。他将图（B）中的细菌描述成“先向前游泳之后再后退”（1692）。（B）真核生物团藻（1700）
（经 John Innes Foundation 许可）

在这一章中，我们将简要回顾研究细胞及其组分，特别是蛋白质、DNA 和 RNA 的主要方法。我们将关注如何将不同类型的细胞从组织中分离出来，在体外培养，如何破碎细胞，分离出纯的细胞器和大分子组分。之后，我们将回顾至今仍在使我们对细胞功能的认识发生着变革的 DNA 重组技术。

这一章是一座桥梁，帮助读者从细胞和分子生物学基础知识过渡到对于大分子如何进行组织和如何共同行使协调细胞和生物体生长、发育的功能的详细讨论。这里所介绍的技术和

方法曾经促成了本书中从始至终谈论的科学发现，今天仍然在被数以万计的科学家们日复一日地使用着。

细胞分离和培养

尽管细胞中的细胞器和大分子能在显微镜下观察到，但是要了解这些组分如何行使功能还需进行进一步的生物化学分析。大多数生物化学方法都需要我们得到大量的细胞，把它们进行物理破碎以分离其中的组分。如果样品是一块由多种细胞组成的组织，不同种类的细胞混在一起。为了尽可能地从一种类型的细胞中获得更多的信息，生物学家想出办法把细胞从组织中分离出来，并把不同类型的细胞分开。对通过这些操作得到的相对同类的细胞群可以直接进行分析，也可以对纯品系进行扩增后再分析。

从组织悬液中分离不同类型的细胞

从含有不同类型细胞的组织中分离单一类型的细胞首先要破坏使细胞连接在一起的细胞间质。通常最好的细胞分离的多类型混合物是从胚胎或新生幼体组织中得到的。组织样品一般要经过蛋白水解酶（如胰蛋白酶和胶原蛋白酶）和化学试剂（如乙二胺四乙酸，EDTA）的处理。蛋白酶可以消化胞外基质中的蛋白质，化学试剂可以螯合细胞连接所依赖的 Ca^{2+}。这样经过温和的搅拌，组织就可以分离为单独的活细胞。

多种方法被用于从混合细胞悬液中分离不同类型的细胞。例如，通过离心，大细胞可以与小细胞分开，密度大的细胞可以与密度小的细胞分开。这些技术最初被用于细胞器和大分子的分离，将在下面介绍。另一种方法基于某些细胞倾向于玻璃或塑料上的特性，利用这种特性可以将它们与黏附性较弱的细胞分开。

对最后这项技术的一个重要改进是基于抗体的特异性结合。可以把特异性地与组织中某一种细胞地表面结合的抗体与各种基质（如胶原质、多聚糖颗粒或塑料）相连，从而形成只有被抗体识别的细胞才能附着的亲和表面。然后，结合的细胞可以通过轻摇，用胰蛋白酶处理，介导黏附的蛋白质或者在基质可降解的情况下（如胶原质）用酶降解基质使细胞脱落下来。

最复杂的细胞分离技术之一是用携有荧光染料的抗体标记特定细胞。被标记的细胞随后可通过荧光激活的电子细胞筛选器与未标记的细胞分离开。在这种先进的仪器中，单个细胞在细流中排成一列，依次通过一束激光，细胞上的荧光被迅速检测到。振荡喷管产生极小的液滴，大多数液滴只含有一个细胞或不含细胞。含有单个细胞的液滴在产生时被自动带上正电荷或负电荷，电性取决于细胞是否带荧光。此后它们在强电场中被分选到适当的容器中。偶然聚集的细胞产生超常的散射，因此没有被带上电荷，从而被分选到盛废液的容器中。这种仪器可以准确地从 1000 个未标记的细胞中分选出 1 个荧光标记的细胞，并且每秒钟能够分选数千个细胞（图 8-2）。

选定的细胞也可以通过从用于镜检的组织切片上进行精细切割得到（讨论见第 9 章）。一种方法是将组织切片被包上一层塑料薄膜，含有目的细胞的区域用红外线激光仪发出的聚焦脉冲照射。这种弱脉冲可以熔化薄膜，使细胞结合在底面上。可以分离这些获得的细胞用于进一步分析。这种技术叫做“激光获取显微切割”，可以用于从肿瘤的不同部位分离和分析细胞，比较它们的性质。一种相关的方法是用激光束直接切出一群细胞，并将它们推入适当的容器中用于进一步分析（图 8-3）。

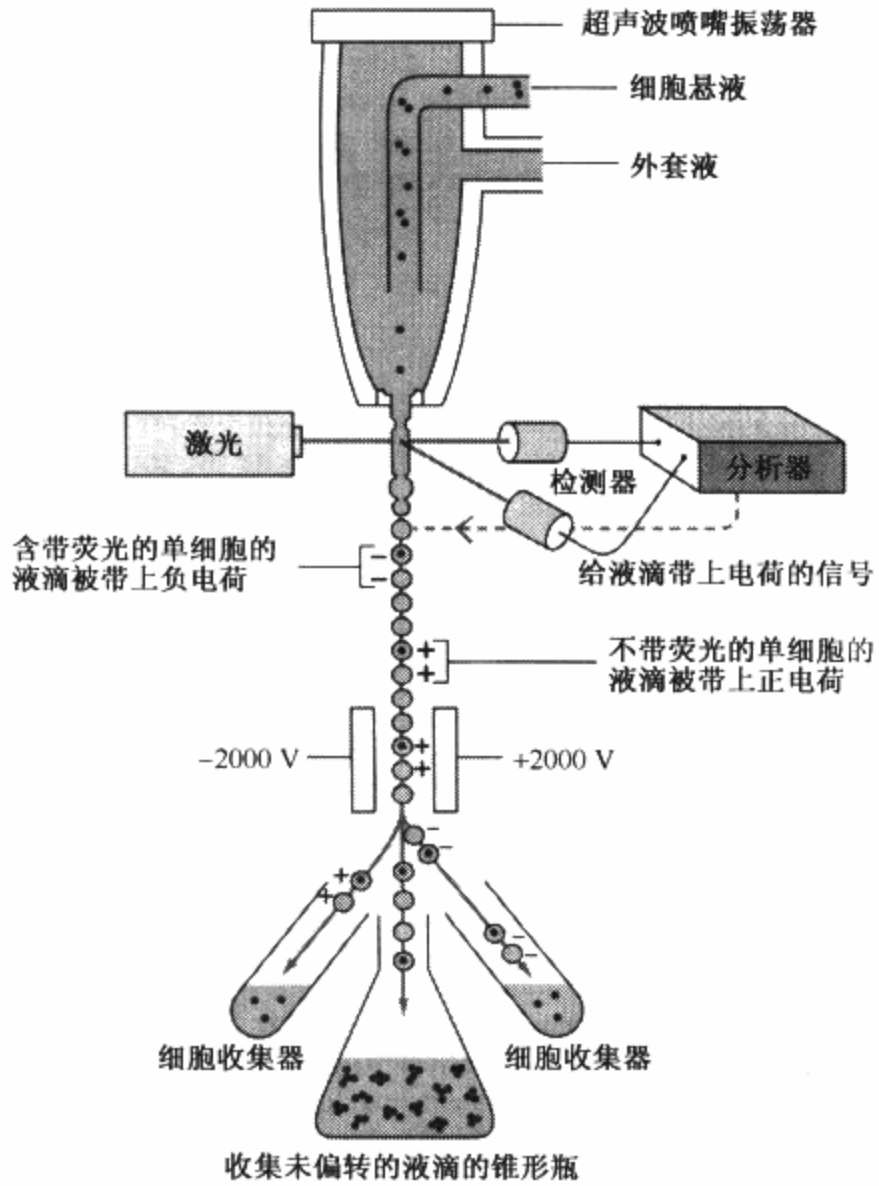


图 8-2　荧光激活的细胞分选仪

细胞通过激光束，引其荧光标记而受到操纵。含有单细胞的液滴根据所含细胞是否带荧光而被带上负电荷或正电荷。之后根据带电性的不同，液滴在电场中被分选到不同的收集管中。注意细胞浓度必须调整合适，以保证不含细胞和含有多细胞聚集体的液滴被分选到废液缸中

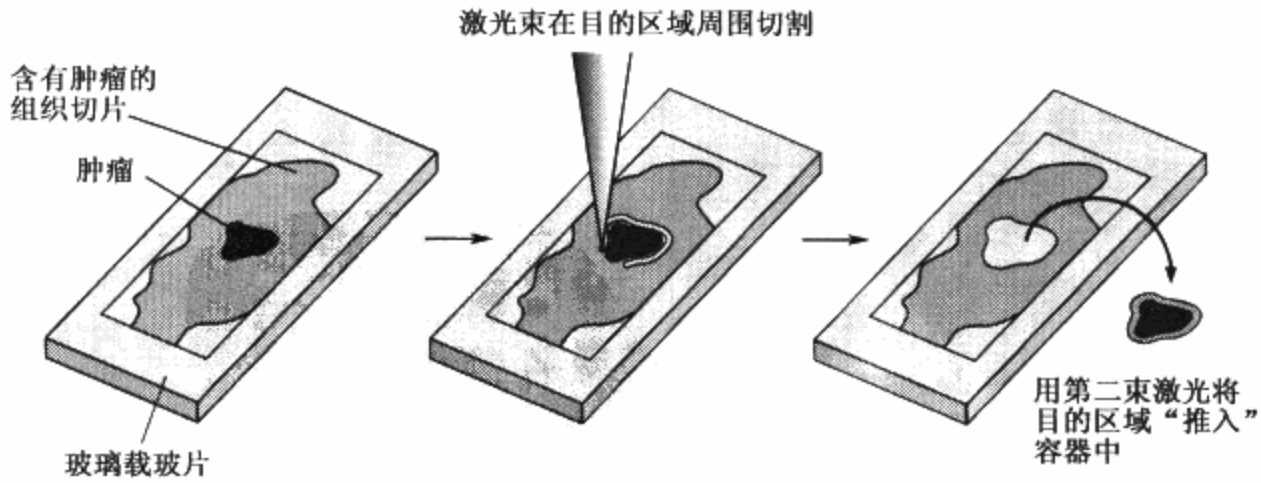


图 8-3　用显微切割技术可以从组织切片中分离目的细胞

这种方法是用激光束切出目的区域，将其推入容器中。这种方法甚至可以从组织样品中分离单个细胞

单一细胞群一旦获得（用显微切割或者任何上面讲到的分离手段）就可以用于生物化学分析。异质的细胞样品也可以用于细胞培养，能使细胞数量得到极大扩增，并可以在严格限

制生长条件的细胞培养皿中研究它们的复杂行为。

细胞可以在培养皿中培养

在合适的环境中，大多数植物和动物细胞都能够生存、复制，甚至可以在组织培养皿中表现出各种分化的性质。可以在显微镜下连续观察细胞，或者分析它们的生物化学性质，研究加入或除去像激素、生长因子这样的小分子对它们产生的影响。除此之外，通过混合两种细胞，可以研究它们的相互作用。有时候，在培养的细胞上进行的实验叫做体外实验（*in vitro*，字面意思是“在玻璃仪器中”），以便与在完整生物体中进行的体内实验（*in vivo*，字面意思是“在活的生物体内”）相区别。但是，因为生物化学家这些术语表达完全不同的意思，所以很容易混淆。在生物化学实验室中，*in vitro* 指的是在非细胞体系的试管中进行的反应，而 *in vivo* 指的是在活细胞中发生的反应（哪怕是培养的细胞）。

组织培养开始于 1907 年一个为解决神经生物学的一场争论而进行的实验，为的是检验叫做神经元学说的假说。神经元学说认为每一条神经纤维都是一个单个神经细胞长出来的，而并非多个细胞融合而成的。为了检验这个观点，人们把脊髓小块置于温暖、潮湿的小室中的

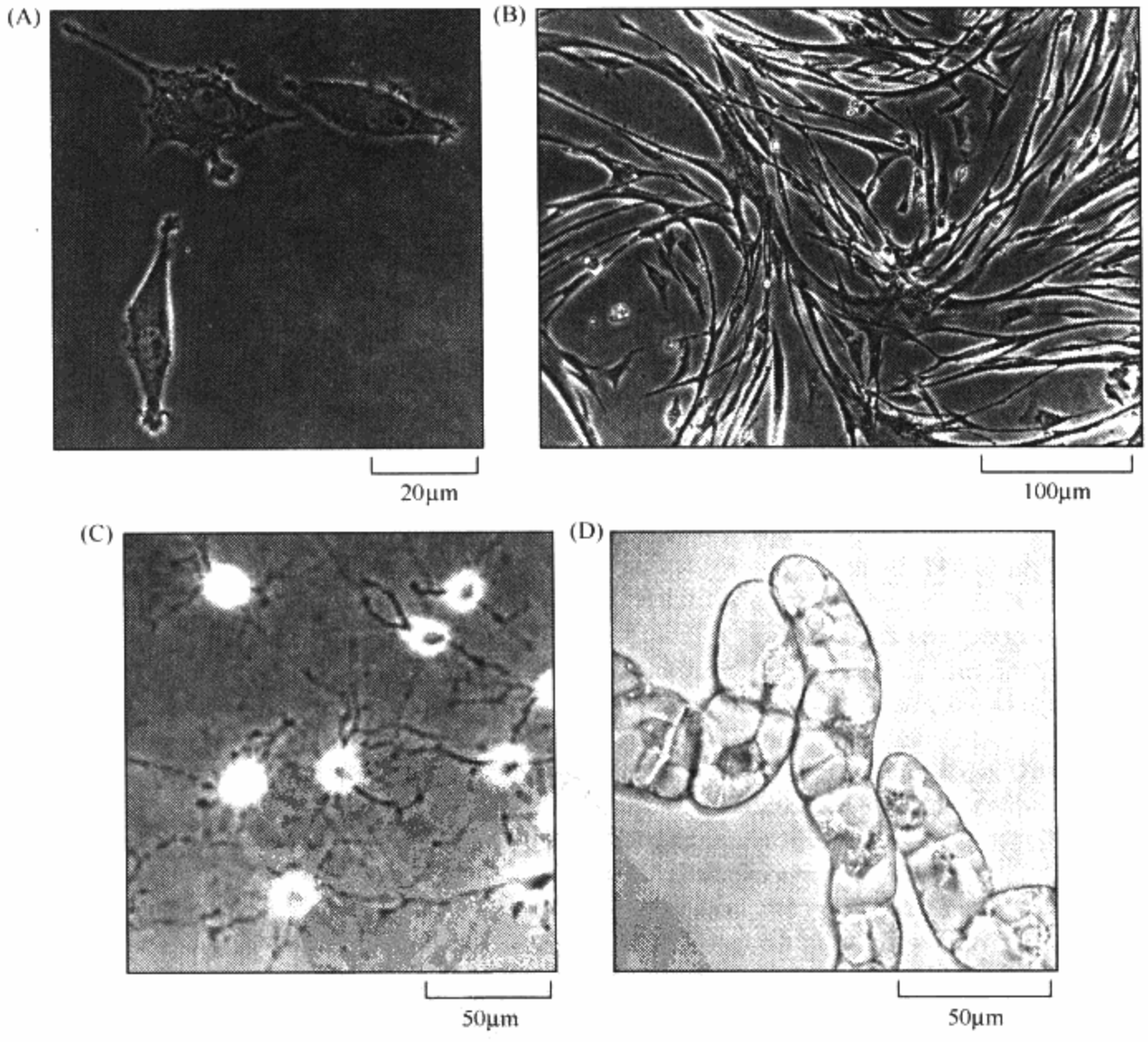


图 8-4　培养的细胞

（A）培养的成纤维细胞的相差显微照片。（B）从培养的成肌细胞显微照片上可以看到细胞在融合成多核的肌肉细胞。（C）培养的少突细胞前体细胞。（D）液体培养基中的烟草细胞，从快速生长的永生细胞系 BY2 培养而来［（A）经 Daniel Zicha 许可；（B）经 Rosalind Zalin 许可；（C）摘自 D. G. Tang et al.，*Cell Biol*. 148：971～984，2000，洛克菲勒大学出版社；（D）经 Gethin Roberts 许可］

组织液凝块上，按照固定的时间间隔在显微镜下观察。过了一天左右，可以看到单个神经细胞伸出细长的纤维，伸入凝块中。于是，神经元学术得到了有力的支持，细胞培养的基础也由此建立。

神经纤维的早期实验所用的小块组织碎片叫做外植体。今天，细胞培养更多是从用前面所介绍的方法进行组织破碎获得的细胞悬液培养而来的。与细菌不同，大多数组织细胞不能适应悬浮生活，需要有固体表面的支持才能进行生长和分裂。对于细胞培养来说，这种支持通常由塑料组织培养皿的表面提供。不过，不同细胞的要求有所不同，许多细胞如果培养皿不包括特定的细胞间质组分（如胶原蛋白或层粘连蛋白），就不能生长和分化。

从一个生物体的组织中直接制备的细胞培养，也就是说没有经过体外增殖，叫做初级细胞培养。这种细胞培养可以经过或不经过不同种类细胞的分选步骤而获得。在大多数情况下，初级细胞培养可以从培养皿中转移进行大量的扩增，得到所谓的次级细胞培养，可以连续培养数周或数月。这些细胞常常表现出与它们的来源相符的许多分化的特征：成纤维细胞不断分泌胶原蛋白；胚胎骨骼肌来源的细胞融合，形成肌纤维，能在培养皿中自发收缩；神经细胞生长出具电敏感性的轴突，并与其他神经细胞建立突触联系；上皮细胞形成大量具有很多完整表皮形状的片层（图 8-4）。因为这些现象发生在培养过程中，所以可以对它们进行对完整的组织无法进行的研究。

在无血清，化学成分已知培养基中鉴别特定生长因子

直到 19 世纪 70 年代早期，组织培养看起来还像是科学和巫术的混合。尽管液体凝块已被含有特定含量的食盐、葡萄糖、氨基酸和维他命等小分子的液体培养基所取代，但大多数培养基中仍然或是含有没有很好地被鉴定的大分子混合物，如马或胎牛血清或鸡胚胎的粗提物。这些培养基今天仍然在常规的细胞培养中使用（表 8-1），但是它们使得研究者很难弄清究竟是哪一种大分子成分是某种细胞生存和正常行使功能所需要的。

表 8-1　一个用于哺乳动物细胞培养的标准培养基的组成

氨基酸	维生素	盐	其他成分	蛋白质（用于非血清化学培养基）
精氨酸	生物素	NaCl	葡萄糖	胰岛素
胱氨酸	胆碱	KCl	青霉素	铁传递蛋白
谷氨酸	叶酸	NaH_2PO_4	链霉素	特异生长因子
组氨酸	尼克酰胺	$NaHCO_3$	酚红	
异亮氨酸	泛酸盐	$CaCl_2$	全血清	
亮氨酸	吡哆醛	$MgCl_2$		
赖氨酸	硫胺			
甲硫氨酸	核黄素			
苯丙胺酸				
苏氨酸				
色氨酸				
酪氨酸				
缬氨酸				

葡萄糖的使用浓度为 5～10mmol/L。除了一两个例外，氨基酸都是 L 型的，使用浓度为 0.1mmol/L 或 0.2mmol/L；维生素的使用浓度低 100 倍，即大约 1μmol/L。血清通常取自马或小牛，加入量为总体积的 10%。青霉素和链霉素是抗生素，用于抑制细菌的生长。酚红是 pH 指示剂，通过控制它的颜色可以将 pH 控制在 7.4。

细胞通常培养在具有合适的附着面的塑料或玻璃容器中。容器保存在 37℃，5%的 CO_2 和 95%的空气中。

这个困难促成了各种无血清化学培养基的发展。除了常规的小分子之外，这些化学成分已知的培养基含有一种或多种细胞生存和增殖所需要的蛋白质。加入的这些蛋白质包括促进细胞增殖的生长因子，把铁送到细胞的铁转移蛋白。许多对特定细胞类型的生长、发育和增殖必不可少的细胞外的蛋白质信号分子是在寻找使细胞正常生活所需的最少物质条件的过程中发现的。因此，化学成分已知的培养基使得寻找新的信号分子的工作更加容易了。

真核细胞系是均质细胞的重要来源

大多数脊椎动物的细胞在培养基中分裂一定的次数之后就会停止，这个过程叫做细胞衰老（讨论见第 17 章）。例如，正常的人类成纤维细胞通常在培养基中分裂 25～40 代就会停止。在这些细胞中，有限的分裂次数是细胞端粒不断变短的结果。细胞端粒指的是盖在细胞的每条染色体末端的重复性 DNA 片段及其结合蛋白质（讨论见第 5 章）。在人类的体细胞中，由于起着保持端粒的作用的端粒酶的失活，端粒在每一次分裂中都会缩短。如果为人类成纤维细胞提供编码端粒酶催化亚基的基因，它们将永远分裂下去，成为永生的**细胞系**。

不过有些人类细胞不是通过这种方式获得永生的。尽管它们的端粒保持了长度，它们仍然会在有限次的分裂之后停止下来，这是因为它们在细胞周期控制点机制中被停滞了（讨论见第 17 章）。为了使这些细胞永生，人们需要提供比端粒酶更多的东西。人们必须通过介导某些来自肿瘤细胞的癌症激活因子使控制点机制失去活性（讨论见第 23 章）。与人类细胞不同，大多数啮齿类动物细胞的端粒酶在分裂中不失活，因此它们的端粒在分裂中不变短。此外，啮齿类动物的细胞在培养过程中可以通过遗传性质的改变使控制点机制失活，自发产生永生的细胞系。

从癌细胞中获得细胞系通常是最简单的办法，但是这样产生的细胞在很多方面都与正常的细胞不同。例如，癌细胞在生长中不需要附着在表面上，可以在培养基中生长达到高得多的浓度。通过病毒或化学物质的改造也可以使正常细胞获得相似的性质。相应地，这样产生的细胞系，再注入易感动物的体内常常会引起肿瘤。改造的和永生的细胞系在细胞研究中都非常有用，它们提供了大量单一类型的细胞，并且可以无限期地保存在－196℃的液氮中，一旦解冻又能恢复活性。但是必须明确的是，这两种细胞系与它们组织中的祖先相比有着极大的不同。一些常用的细胞系在表 8-2 中列出。

表 8-2　一些常用的细胞系

细胞系*	细胞类型和来源
3T3	成纤维细胞（小鼠）
BHK21	成纤维细胞（叙利亚仓鼠）
MDCK	上皮细胞（狗）
HeLa	上皮细胞（人）
PtK1	上皮细胞（跳鼠）
L6	成肌细胞（家鼠）
PC12	嗜铬细胞（家鼠）
SP2	浆细胞（小鼠）
COS	肾细胞（猴子）
293	肾细胞（人）；经过腺病毒转化
CHO	卵巢细胞（中国仓鼠）
DT40	用于有效再结合的淋巴瘤细胞（鸡）
R1	胚胎干细胞（小鼠）
E14.1	胚胎干细胞（小鼠）
H1、H9	胚胎干细胞（人）
S2	类巨嗜细胞（果蝇）
BY2	未分化的分生组织细胞（烟草）

* 这些细胞系中的很多来自于肿瘤。它们都可以在培养基中无止境地分裂，至少能部分表达它们祖先的特定性状。BH21 细胞、HeLa 细胞和 SP2 细胞能在悬液中高效生长，而其他大多需要固体的培养基底质才能繁殖

在医学上最有前景的细胞系是人类的胚胎干细胞（ES）。这些细胞是从人类早期胚胎的内部细胞团中获得的，可以永久地增殖并能生成身体的任一部分（讨论见第 21 章）。胚胎干细胞可以取代或修复损伤或病变的组织，其潜能会使医学产生重大的变革。

尽管一个细胞系中的细胞非常相似，它们并非完全相同。细胞系的遗传一致性可以通过细胞克隆得到改进，即分离单个细胞让其繁殖成大细胞群体。在这样的细胞群或称克隆中，所有的细胞都是一个共同祖先的后代。细胞克隆最大的用途之一是分离具有特异基因缺陷的细胞系。通过研究特定蛋白质缺陷的细胞可以了解这种蛋白质在正常细胞中的功能。

一些细胞培养中的重大进展在表 8-3 中列出。

表 8-3　组织和细胞培养发展中的一些重要事件

1885 年	Roux 指出鸡胚胎细胞可以在动物体外的盐溶液中存活
1907 年	Harrison 在淋巴结中培养出了两栖动物脊髓，从而证明轴突是单个神经细胞延伸产生的
1910 年	Rous 用鸡肿瘤细胞的提取物滤液诱导出了肿瘤，后来发现其中含有 RNA 病毒（劳氏肉瘤病毒）
1913 年	Carrel 证明细胞可以在定期补充营养的无菌培养基中长期培养
1948 年	Earle 和同事分离出了单个 L 细胞系的细胞，并证明它们在组织培养中形成细胞克隆
1952 年	Gey 和同事从人的宫颈癌细胞中建立了连续的细胞系，后来成为著名的 HeLa 细胞系
1954 年	Levi-Montalcinci 及其合作者证明神经生长因子（NGF）在培养中促进轴突的生长
1955 年	Eagle 首次对细胞培养必需的营养成分进行了系统的研究，发现动物细胞可以在添加了少量血清蛋白的成分确定的小分子混合物培养基中繁殖
1956 年	Puck 及其合作者从 HeLa 细胞培养中筛选对生长条件要求不同的突变体
1958 年	Temin 和 Rubin 发明了一种用纯化的劳氏肉瘤病毒大量感染培养的鸡细胞的方法。在接下来的 10 年中，Srocker、Dulbecco、Green 和其他的病毒学家建立了这种病毒和其他病毒改造细胞的典范
1961 年	Hayflick 和 Moorhesd 表明培养中的人类成纤维细胞在进行有限次数的分裂之后死亡
1964 年	Littlefield 将 HAT 培养基用于杂种体细胞的选择性培养，与细胞融合技术一同，使这些体细胞具有遗传上的易感性 Kato 和 Takeuchi 从胡萝卜的根细胞组织培养得到了整株的胡萝卜。
1965 年	Ham 使用成分确定的无血清培养基克隆培养某些哺乳动物细胞 Harris 和 Watkins 用病毒诱导的人类细胞和小鼠细胞首次造出了哺乳动物异核体
1968 年	Augusti-Tocco 和 Sato 采用小鼠神经细胞瘤（成纤维细胞）进行组织培养，并分离出具有电易感性和使神经进程延长的克隆。一些其他的分化细胞也在这时期分离出来，包括骨骼肌细胞和肝细胞系
1975 年	Köhler 和 Milstein 制造出第一个分泌单克隆抗体的杂交瘤细胞系
1976 年	Sato 及其合作者发表了最初的一系列论文，提出不同的细胞系需要不同的激素和生长因子混合物以在无血清培养基中生长
1977 年	Wigler、Axel 及其合作者在 Graham 和 van der Eb 的方法基础上，发展出一种将单拷贝的哺乳动物基因诱导入培养细胞的有效方法
1986 年	Martin、Evans 及其合作者分离并培养出小鼠多功能胚胎干细胞
1998 年	Thomson、Gearhart 及其合作者分离出人类胚胎干细胞

细胞可以融合形成杂种细胞

可以把一个细胞与另一个细胞融合，形成具有两个独立的核的异核体。通常，把细胞悬液用灭活的病毒或者聚乙二醇处理可以改变质膜的性质，诱导它们融合。例如，无活性的鸡

红细胞核融合后暴露于生长期的组织培养细胞的胞质中时，被激活产生 RNA，并最终进行 DNA 的复制。膜蛋白能够在内质网面上移动（讨论见第 10 章）的第一个直接证据来源于小鼠细胞与人细胞融合的实验：尽管小鼠细胞与人细胞的细胞表面蛋白在开始的时候位于它们的半个细胞所属的一侧，但是它们很快就混合分布在了整个细胞的表面。

最终，异核体进入有丝分裂，两个独立细胞核的核膜解体，所有的染色体合并在一起，产生杂交细胞（图 8-5）。尽管这样的杂合细胞也可以进行克隆产生杂交细胞系，但是它们倾向于丢失染色体，因而在遗传上是不稳定的。由于未知的原因，小鼠-人的杂交细胞系总是倾向于丢失人的染色体。随机的丢失产生了只含有一条或者几条人染色体的多种杂交细胞系。这种现象被应用于在人类基因组中对基因进行定位。例如，只有含有 11 号人类染色体的杂交细胞能够合成胰岛素，说明编码胰岛素的基因位于 11 号染色体上。同样的杂交细胞还可以作为染色体特异性人类 DNA 文库的 DNA 来源。

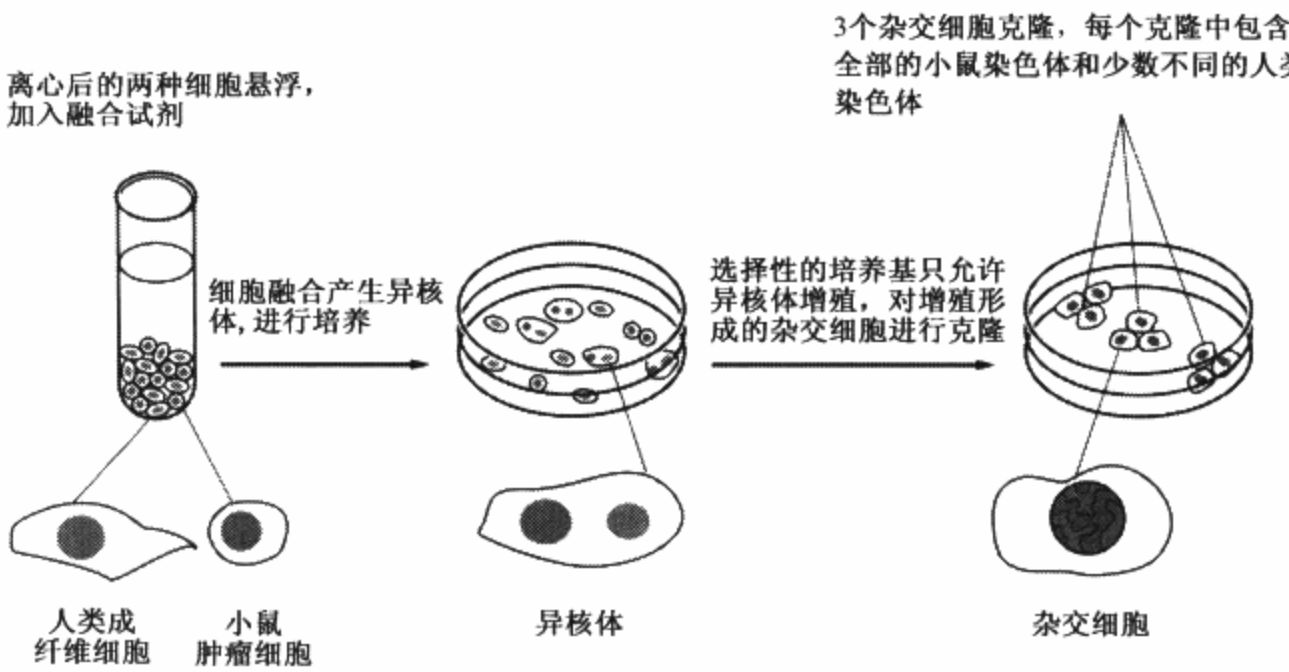


图 8-5　杂交细胞的制备

融合人类和小鼠细胞产生异核体（每个含有两个或更多细胞核），最终成为杂交细胞（每个只含有一个融合的细胞核）。这些特异的杂交细胞对于测定人类基因在哪条染色体上是十分有用的，因为大多数人类染色体以随机的方式迅速丢失了，只留下一条或几条染色体。融合产生的杂交细胞通常能够保留它们多数的染色体

杂交瘤细胞系是单克隆抗体的持久来源

1975 年，一种特殊的杂交细胞的产生使得作为细胞生物学工具的抗体的生产发生了革命。这项技术通过扩增从分泌单一抗体的 B 淋巴细胞得到的细胞克隆可以得到大量的单一抗体。但实际的问题是，B 淋巴细胞通常在细胞培养中只有有限的寿命。为了打破这个限制，人们把从免疫过的小鼠或家鼠中获得的单个 B 淋巴细胞与“永生的”B 淋巴瘤细胞融合在一起，通过选择产生了既能产生特异抗体又能无限制地繁殖的杂交细胞。这些**杂交瘤细胞**被作为一个个细胞克隆，每一个克隆都可以是一种**单克隆抗体**的稳定来源（图 8-6）。这些抗体可以识别单一种类的抗原位点。例如，蛋白质表面特定的五六个氨基酸侧链的丛。一致的特异性使单克隆抗体对于大多数研究比传统的抗血清有用得多，因为抗血清中通常含有识别大分子上不同抗原位点的多种抗体。

杂交瘤技术最大的优点是可以制造复杂混合物中少量组分分子的单克隆抗体。在这种混合物普通的抗血清中，识别这种少量组分的抗体因含量太少而难以使用。但是如果把产生抗

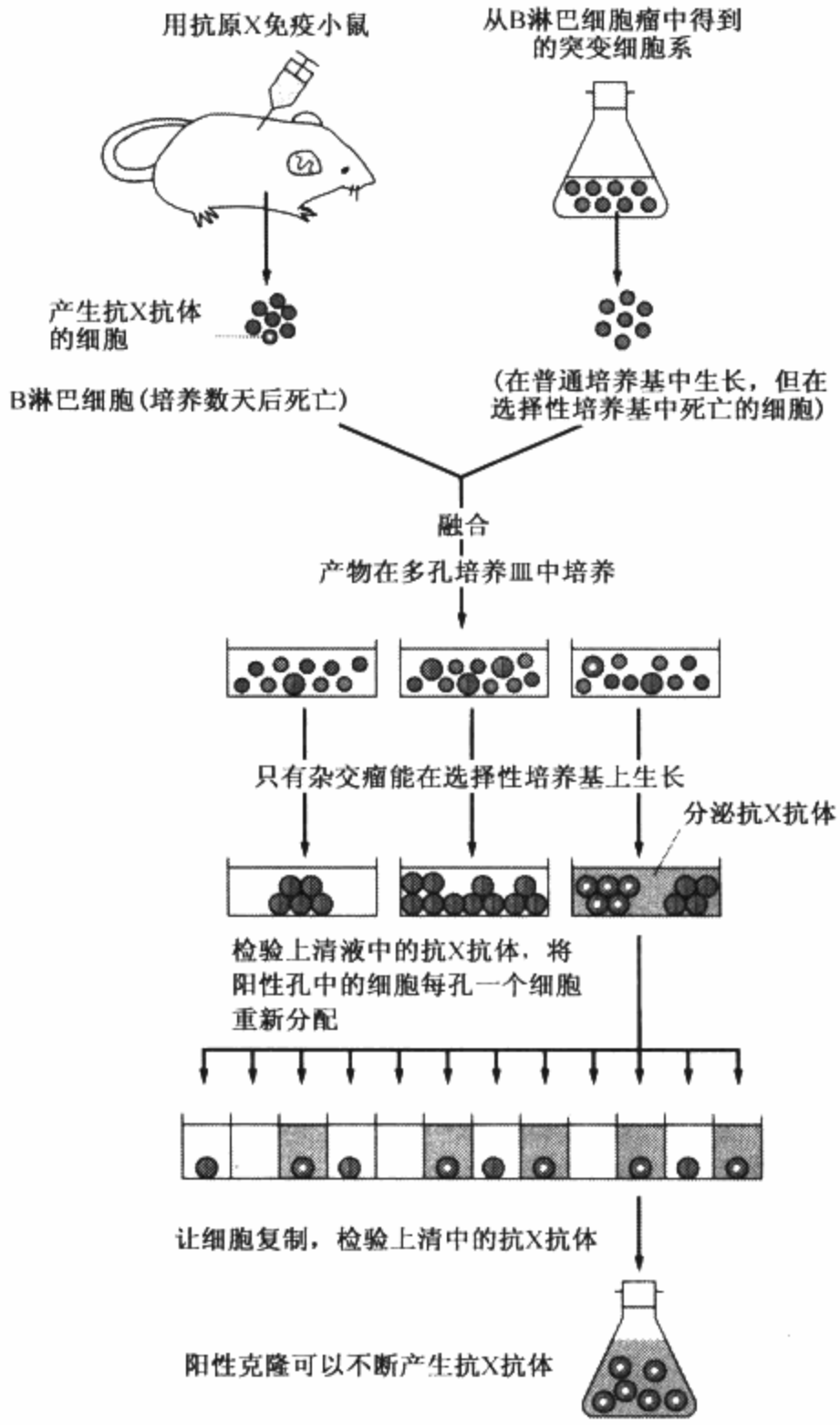


图 8-6　分泌针对某一特异抗原的单克隆抗体的杂交瘤细胞的制备

这里有我们感兴趣的抗原——抗原 X。融合后选用的选择性的培养基中含有能够阻断一般的核苷酸生物合成途径的抑制子（氨基蝶呤）。因此细胞只能通过旁路途径合成核苷酸。在从肿瘤细胞得到的缺陷型细胞系中，该途径是缺陷的，但是在免疫过的小鼠细胞中是完好的。两种用于初始融合的细胞都不能独自存活，只有杂交细胞能存活

血清的各种组分的 B 淋巴细胞做成淋巴瘤，就可以对混合体系中的单个淋巴瘤克隆进行鉴别，筛选出产生所需单克隆抗体的一类，进行持续的扩增，获得无限量的抗体。所以，在理论上，生物样品中任一种蛋白质的单克隆抗体都可以得到。

抗体一旦获得，可以用作特异性探针，既可以用于追踪和对其抗原蛋白进行定位，也可以用于纯化蛋白质以研究其结构和功能。因为迄今为止，一个普通哺乳动物细胞中预计的 10 000～20 000 种蛋白质中只有一小部分被分离出来了，许多从破碎的细胞提取物得到不纯的蛋白质混合物，它们的单克隆抗体鉴定出了新的蛋白质。随着单克隆抗体和我们将简短介绍的快速的蛋白质鉴定方法的应用，新蛋白质和基因的鉴定和定性不再是难题，而取而代之

的主要问题将是通过一系列我们将在这章最后讨论的强有力的工具判定这些蛋白质的功能。

小结

可以把组织分离成细胞，从中纯化出单一的细胞种类，用于生化分析或进行细胞培养。很多动物和植物细胞当给它们提供含有营养和特异蛋白质生长因子的合适的培养基时，都能够生存和增殖。尽管大多数动物细胞会在有限次分裂之后死亡，能在培养基中自发繁殖的永生细胞（或者通过基因操作加入某些基因产生的细胞）可以作为细胞系永久保存。从单个细胞祖先可以得到的细胞克隆，用这种方式能够获得单个蛋白质缺陷的均一突变细胞群。两个细胞可以融合成含有双核的异核体，用于研究原来分属两个细胞的细胞组分相互作用。异核体最终形成核融合的杂交细胞。因为这些细胞会丢失染色体，它们提供了把特定基因定位在染色体上的捷径。杂交细胞的一种——杂交瘤细胞，被广泛用于无限量地生产可以探测和纯化细胞蛋白质的单克隆抗体。

细胞组分的分离

尽管生物化学分析要求破坏细胞的结构，采用温和的分离技术仍然可以在分离各种细胞组分的同时保持它们各自的功能。正如可以把组织分离成不同类型的活细胞一样，细胞也可以分离成功能细胞器和大分子。在这一节中我们将讨论对细胞器和蛋白质进行纯化和生物化学分析的方法。

可以通过超离心分离细胞器和大分子

有多种方法可以打破细胞：渗压震扰或者超声波振荡，通过小孔或者进行搅拌。这些方法可以将细胞的许多膜结构（包括质膜和内质网膜）破碎成小片，之后立即又围合成封闭的小泡。如果温和地破碎，可以保持细胞核、线粒体、高尔基体、溶酶体和过氧化物酶体等细胞器的完整。这样，细胞的悬液变成了黏浆（叫做匀浆或提取物），其中含有多种膜包被的细胞器，并且每一种具有特定的大小、电荷和密度。如果对匀浆方法进行认真选择的话（对每种细胞器都进行反复试验），那么各种组分（包括内质网形成的小泡，叫做微体）中的大多数都能保持原来的生物化学特性。

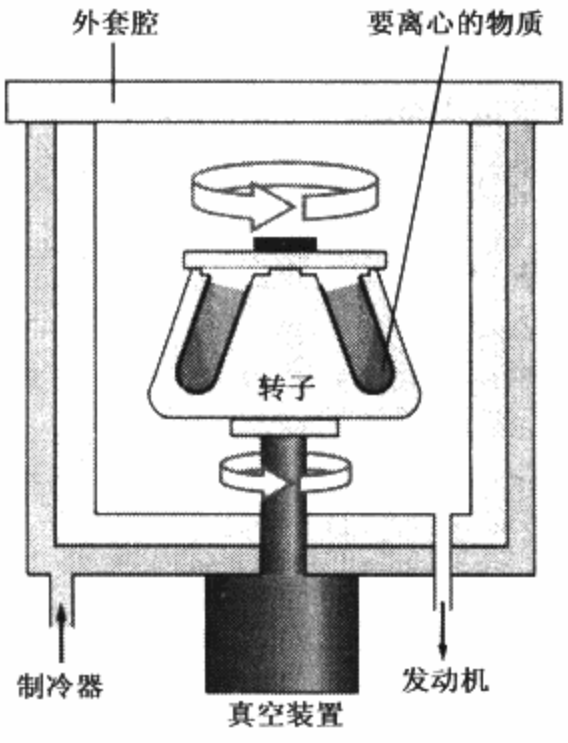


图 8-7　制备超速离心

将含有样品的离心管插入金属转子环形排列的柱状孔中。高速的离心产生强大的离心力，使样品粒子沉降下来。真空减小了摩擦，防止了转子生热，保证制冷系统把温度控制在 4℃

匀浆中的不同细胞组分需要进行分离。直到 20 世纪 40 年代早期一种名叫制备超速离心机的仪器商业化之后，细胞匀浆可以在其中进行高速旋转，这样的分离才成为可能（图 8-7）。这种方法是通过细胞组分的大小和密度进行分离的：一般来说，最大的组分受到最大的离心力，因而运动得最快。在较低的速度下，像细胞核这样的大块组分会沉降在离心管底；在稍高一点的速度下，线粒体会沉降；在更高的速度下，经过更长的离心时间，闭合的小泡和核糖体相继沉降下来（图 8-8）。所有这些组分都不纯，但是大多数的杂

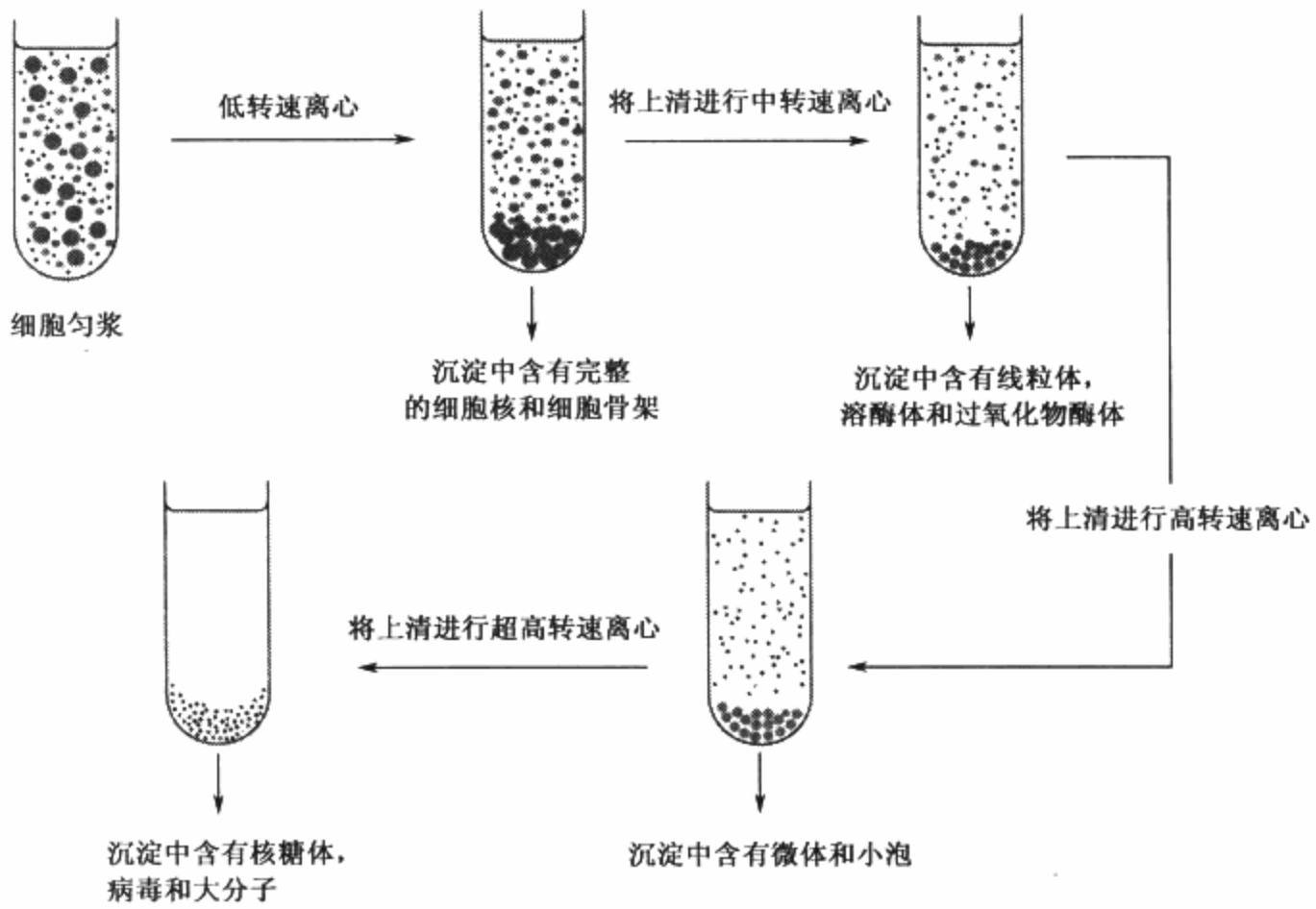


图 8-8　通过离心进行细胞组分的分离

不断增加转速反复离心可以分离细胞匀浆中的各种组分。通常，亚细胞组分越小，沉淀所需的离心力越大。各个离心步骤的一般值是：低转速 1000 倍重力加速度，10min；中转速 20 000 倍重力加速度，20min；高转速 80 000 倍重力加速度，1h；超高转速 150 000 倍重力加速度，3h

质可以通过重新悬浮沉淀多次离心去除。

离心是分离的第一个步骤，但它只能分离大小相差很大的组分。通过将匀浆薄层加在充满可溶性盐溶液的离心管上层，得到更好的分离效果。在离心过程中，混合物中的不同组分以不同的速率沉降，在盐溶液中形成不同的条带，这个过程叫做速度沉降［图 8-9(A)］。为了使流程能够有效进行，要防止对流导致的条带混合，通常当密度大的溶液（如含有细胞器的溶液）位于密度小的溶液（如盐溶液）之上时，对流就会发生。这个问题是用一种特殊的混合方法解决的，可以在离心管中注满建立了粗略密度梯度的蔗糖溶液。因为产生的梯度场中，每一层都比上层的密度大，所以不会因为对流而破坏建立的梯度场。

当在这样的蔗糖密度梯度场中沉降时，不同的细胞组分分离到了不同的条带中，可以被分别收集。每一个组分沉降的相对速率主要由其大小和性状决定，通常用沉降系数或 S 值表示。今天的超高速离心可以达到 80 000r/min 的转速，产生高达 500 000 倍重力的离心力。在这样强大的力的作用下，即使是像 tRNA 和简单的酶这样小的大分子，也能以相当的速度沉降下来，根据大小彼此分离开。沉降系数的测量通常有助于判断从细胞中得到的大分子物质的大小和规则装配体的亚基组成。

超速离心也基于细胞组分的浮力密度进行分离，而与它们的大小和形状无关。在这种情况下，细胞样品通常要在含有极高浓度的蔗糖或氯化铯的陡峭的密度梯度场中进行沉降。每一种细胞组分以图 8-9(A) 中的方式进行沉降，但是最终到达与其密度相等的位置。组分在这个位置悬浮不能继续移动。因此离心管中产生了一系列不同的条带，最接近离心管底部的条带含有浮力密度最大的细胞组分［图 8-9(B)］。这种方法叫做平衡沉降。这种方法非常灵

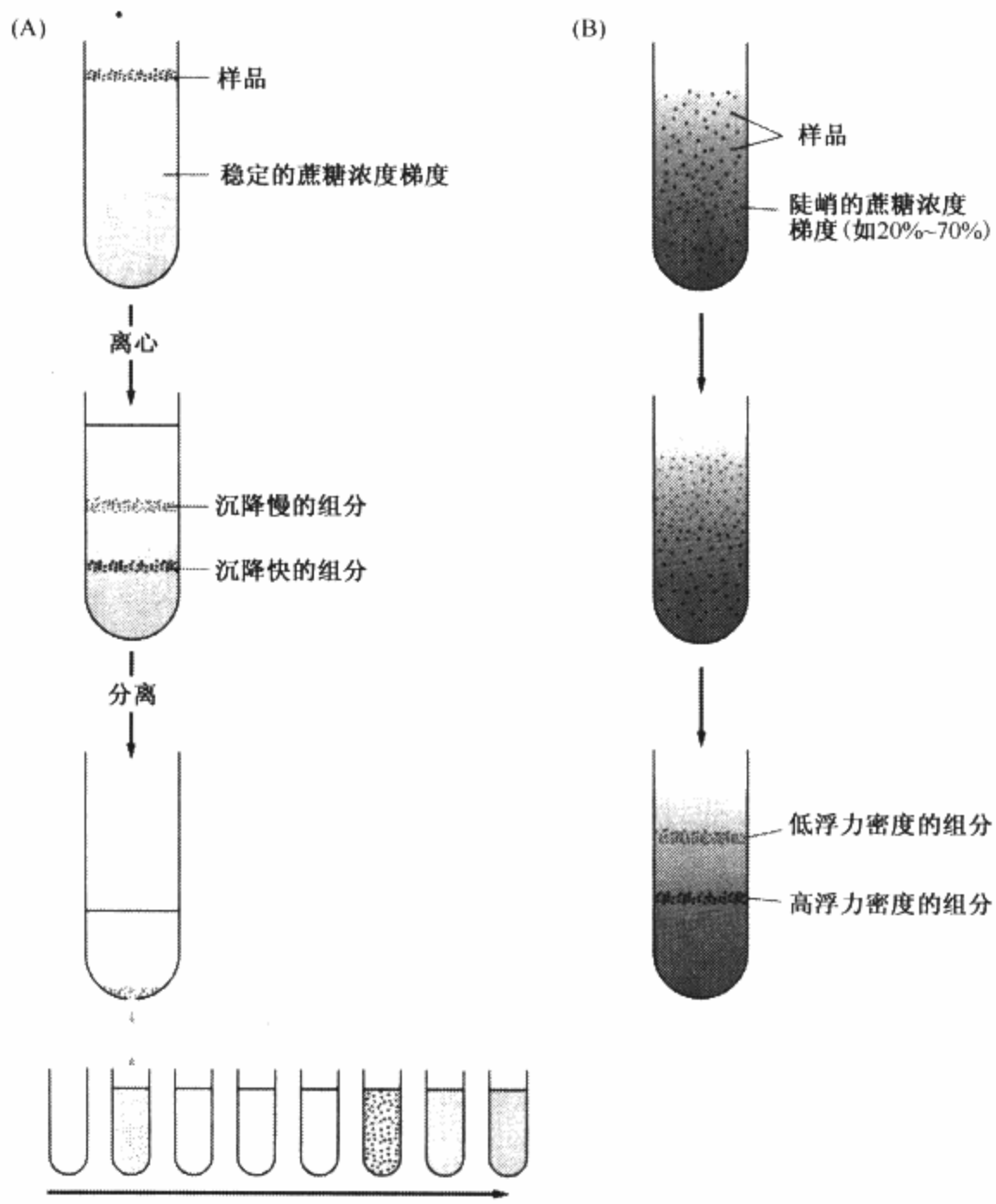


图 8-9　速度沉降和平衡沉降的比较

在速度沉降（A）中，加在蔗糖溶液上层的亚细胞组分根据大小和形状的不同在不同的转速沉淀下来。为了控制由微小的温度或浓度差异导致的条带间对流混溶，自上而下建立浓度连续递增的蔗糖浓度梯度（通常是浓度从5%到20%的蔗糖溶液）。通过离心，不同组分分别被收集，常常采用图中所示的离心管穿孔和液滴收集的方法。在平衡沉降（B）中，细胞组分不断下降，直到到达与其密度一致的位置。尽管在这里展示的是蔗糖密度梯度，但是在蛋白质和核酸的分离中特别有用的更大的浓度梯度可以通过氯化铯获得。最后位于平衡未知的条带可以进行收集

敏，甚至可以用来分离含有同位素的大分子混合物，如可以把含有^{13}C或^{15}N的大分子与含有较轻同位素（^{12}C或^{14}N）的大分子分离开。实际上，氯化铯的方法发明于1957年，用于从暴露于含有核苷酸前体的生长期菌落产生的DNA中，把标记的DNA与未标记的DNA分离开。这个经典实验提供了DNA半保留复制的直接证据（图5-5）。

在非细胞体系中可以解释的复杂细胞过程的分子细节

对通过超速离心得到的细胞器和其他大的亚细胞组分的研究对我们认识不同细胞组分的功能做出了巨大的贡献。例如，用离心纯化出的线粒体和叶绿体进行的实验证明了这些细胞器在把能量转化为细胞可利用形式的过程中发挥着主要作用。与之相似，粗面和光面内质网形成释放的囊泡（微体）被作为这些完好的细胞的组成部分的模型进行了分离和分析。

这种手段的延伸，通过建立非细胞体系，排除了在活细胞中发生的所有复杂的副反应，使得对于许多其他生物过程的研究成为可能。在这种情况下，要分离细胞匀浆以纯化其中对于催化所感兴趣的过程所需的每一种大分子。例如，为了研究蛋白质的合成机制，人们首先对能够把 RNA 分子翻译蛋白质的细胞匀浆进行了分离，逐步得到了核糖体、tRNA 和共同构成蛋白质合成机器的各种酶。一旦得到了纯组分，就可以单独加入或抑制每一种组分，以判定其在全过程中的具体作用。今天的一个主要目标是在提纯的非细胞体系中重建每一个生物过程，以鉴定组分和功能机制。表 8-4 中列出了这一重大研究领域中的一些重要事件。

表 8-4　非细胞体系发展中的一些主要事件

1897 年	Buchner 指出酵母的非细胞提取物可以使糖发酵产生二氧化碳和酒精，奠定了酶学的基础
1926 年	Svedberg 发明了第一台分析超速离心机，并用它估测了血红蛋白的分子质量为 68 000Da
1935 年	Pickels 和 Beams 给离心机增添了多种新性能，使之成为一种制备仪器
1938 年	Behrens 采用差速离心的方法分离肝细胞的细胞核和细胞质，Claude、Brachet 和 Hogeboom 等人在 20 世纪 40 年代和 50 年代初进一步改进了这种方法，用于细胞器的分离
1939 年	Hill 指出受光照的叶绿体可以进行光合作用
1949 年	Szent-Györgyi 指出从骨骼肌细胞中分离的肌原纤维在加入 ATP 时可以收缩。1955 年 Hofmann-Berling 建成了类似的非细胞体系，纤毛可以在其中打动
1951 年	Brakke 用蔗糖密度梯度离心的方法分离一种植物病毒。
1954 年	de Duve 用离心方法相继分离了溶酶体和过氧化体。
1954 年	Zamecnik 和同事创造了蛋白质合成的第一个非细胞体系。在其后的 10 年中，遗传密码被弄清
1957 年	Meselson、Stahl 和 Vinograd 发明了氯化铯平衡密度梯度离心，用于分离核酸
1975 年	Dobberstein 和 Blobel 在非细胞体系中展示了蛋白质跨膜转运
1976 年	Neher 和 Sakmann 发明了测定单个离子通道活性的片箝技术
1983 年	Lohka 和 Masui 从在体外进行整个细胞周期活动的蛙卵中得到了浓缩提取物
1984 年	Rothman 和同事在非细胞体系中重建了体外高尔基小泡转运

可以通过色谱分离蛋白质

分离蛋白质最常用的方法是用柱状色谱，让蛋白质混合物通过装有多孔固体介质的色谱柱。不同蛋白质依据它们与介质的相互作用，发生了不同程度的阻滞，在依次流出柱底时可以进行收集（图 8-10）。通过介质的选择，可以根据蛋白质的电荷（离子交换色谱）、疏水性（疏水色谱）、大小（分子筛色谱）以及与特定小分子或其他大分子的亲和性（亲和色谱）来分离蛋白质。

很多种介质都已经商业化了（图 8-11）。离子交换色谱柱中装有带有正电荷或负电荷的小珠，因而蛋白质可根据其表面电荷的分布而分离。疏水色谱柱中装有带突出的疏水侧链的小珠，所以有疏水区暴露的蛋白质受到阻滞。分子筛色谱柱中装有多微孔小珠，可依据蛋白质的分子质量进行分离：在过柱的过程中，能够进入微孔的小分子在小珠内部发生滞留，而大分子始终位于小珠之间的溶剂中，移动得更快，因此先从柱中流出。除了分离小分子之外，分子筛色谱也是测定分子质量的方便手段。

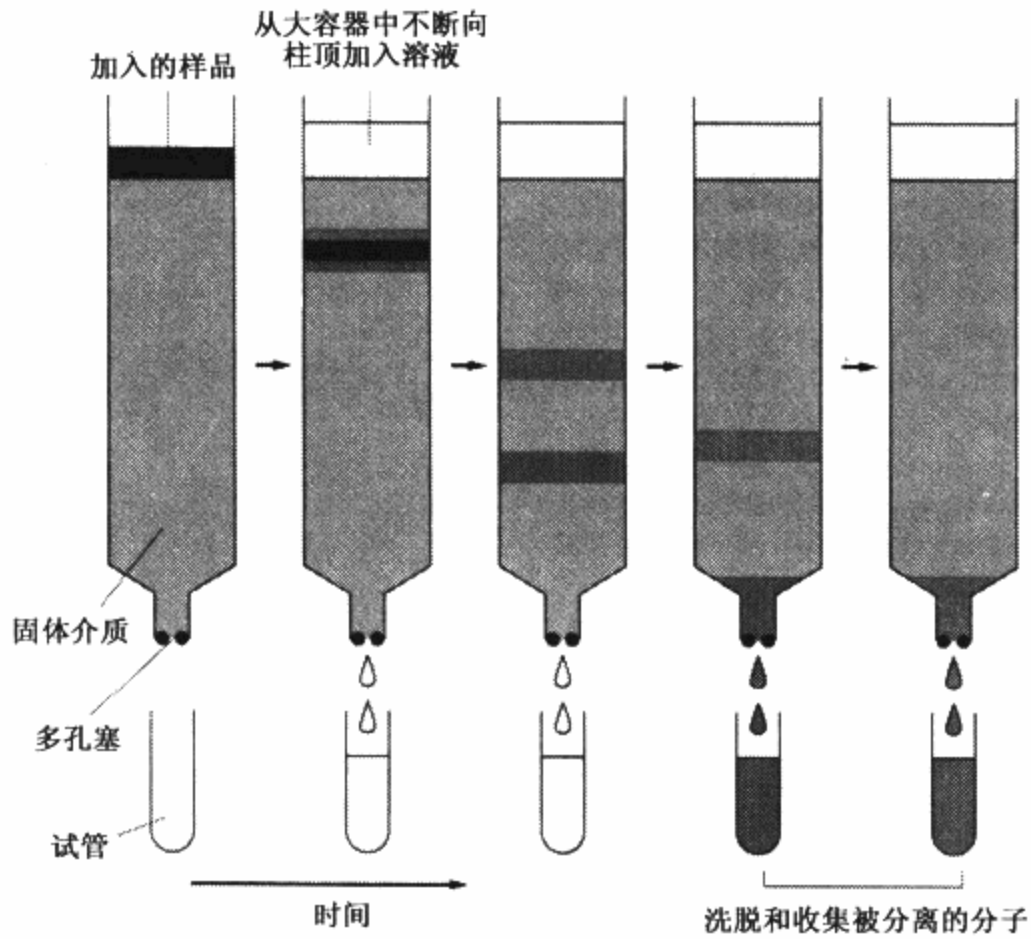


图 8-10　用柱状色谱分离分子

不同分子的混合物样品加在填装有纤维素等可透过的介质的玻璃或塑料柱上方，浸于溶液中。用大量的溶液洗柱，分别收集从柱底留下的液体。因为不同组分通过柱时移动的速度不同，所以它们可以被收集到不同的管中

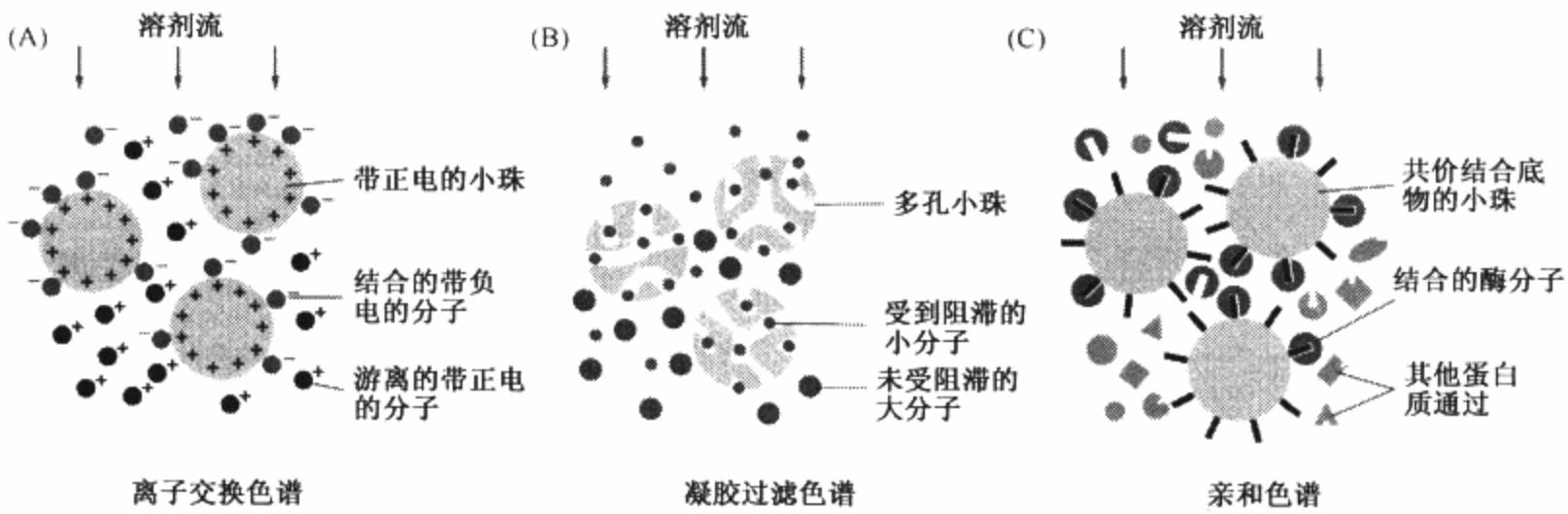


图 8-11　色谱中使用的 3 类介质

在离子交换色谱（A）中，可溶性介质携有离子，可以阻滞带有相反电性的分子。用于分离蛋白质的介质有带正电荷的二乙基氨基乙基纤维素（DEAE-纤维素）、带负电荷的羧甲基纤维素（CM-纤维素）和磷酸化纤维素。基于琼脂糖和其他多聚体的类似介质也常常被采用。溶解的分子与离子交换介质之间的作用强度既与过柱溶液的离子强度有关，也与其 pH 有关，因此可以连续调节以进行有效分离（如图 8-12 所示）。在分子筛色谱（B）中，介质是惰性的多孔物质。能够进入微孔的小分子在小珠内部发生滞留，因此移动得更慢。交联的多糖（葡聚糖、琼脂糖或丙烯酰胺）制成的商业化的小珠有各种不同的孔径，适合于分离不同大小的分子，其分离范围可从不到 500 到高于 5×10^6。亲和色谱（C）共价连接有抗体分子或酶底物等特异性配基的可溶性介质，可与特定的蛋白质结合。与色谱柱上固定的底物结合的酶分子可用含有高浓度游离形式底物的溶液洗脱下来，与固定抗体结合的分子则可以通过用高盐溶液或高、低 pH 溶液解离抗体-抗原复合物洗脱。通过一次亲和色谱通常就可以得到很高的纯度

介质（如纤维素）的不均一性导致了溶剂流的不稳定，从而限制了传统柱状色谱的分辨率。新发明的微球状（直径 3～10μm）的色谱树脂（通常以二氧化硅为主要成分）可以用特殊的设备填装产生均一的柱床。**高效液相色谱（HPLC）**柱可以达到很高的分辨率。因为它们填装得很紧密，所以除非施加高压，HPLC 柱的流速将非常之慢。正因如此，这些柱通常在钢筒中填装，并需要非常精密的泵阀系统以产生能达到大约每分钟一个柱床体积的流速的高压。在传统色谱中，为了给要分离的溶质足够的时间与大个介质颗粒的内部达到平衡，流速必须控制得很低（通常大约每小时一个柱床体积）。而在 HPLC 中溶质很快与微球的内部达到平衡，所以与介质亲和性不同的溶质即使在很高的流速下也能够有效地分离。这使得大多数使用传统色谱即使粗分都需数小时才能完成的分离可在数分钟之内完成。因此，HPLC 成为分离许多蛋白质和小分子的好方法。

亲和色谱利用了蛋白质上的特异性结合位点

如果从蛋白质混合物开始实验，用这些色谱不会得到高纯度的产品：通常每个色谱流程不能达到 20 倍以上的提纯效果。因为大多数蛋白质在细胞总蛋白质中的含量都少于 1/1000，所以通常需要连续使用多种色谱来达到足够的纯度（图 8-12）。**亲和色谱**是一种更有效的手段，它利用了蛋白质表面具有重要生物学意义的结合作用。如果一个底物分子与多糖小珠之类的惰性介质共价结合的话，作用于这种底物的酶常常能够被介质特异性地留住，并在洗脱（洗下）时以近乎纯品的形式被洗出。与之类似，经特别序列构建的寡核苷酸短链可以以这种方式固定，用以纯化通常与染色体上的该序列识别的 DNA 结合蛋白（图 7-30）。也可以将特异抗体连接在介质上用来纯化它们所识别的蛋白质分子。由于亲和色谱的高度特异性，在一个流程中纯度可以提高 1000～10 000 倍。

任何基因都能用下一节中要讨论的 DNA 重组技术修饰，以简单而快速地通过亲和色谱纯化带有特定分子标签的蛋白质，得到产物（图 8-48）。例如，组氨酸序列能与包括镍和铜在内的某些金属离子结合。如果用基因工程技术在一个蛋白质的任一端加上一小段组氨酸残基序列，那么经过轻微修饰的蛋白质可以被固定有镍离子的色谱柱选择性吸附。金属螯合色谱可以用来从复杂的分子混合物中纯化出经过修饰的蛋白质。在有些情况下，可以把整个蛋白质用作分子标签。当把小分子酶谷胱甘肽转移酶（GST）与目标蛋白质相连后，产生的融合蛋白质可以用含有与 GST 特异性紧密结合底物分子谷胱甘肽的亲和柱纯化（图 8-50）。

作为最后这项技术的改进，一段可以被特异蛋白酶切开的氨基酸序列可以设计在目的蛋白质与 GST 标签之间。所采用的蛋白酶的切割位点，例如，在凝血中起作用的 X 因子，在蛋白质中随机偶然出现的概率很小。这样，标签随后可被特异性地从切割位点切掉而不破坏纯化出的蛋白质的结构。

蛋白质的大小和亚基组成可以通过 SDS-聚丙烯酰胺凝胶电泳鉴定

蛋白质通常带有正的或负的净电荷，正负取决于它们所含的带电荷氨基酸残基组成。当对蛋白质溶液外加电场时，蛋白质分子依据它们的净电荷、大小和形状具有一定的前进速度。这项技术叫做电泳，最早用来分离自由水溶液或者如淀粉这样的多孔固体介质支撑下的溶液中的蛋白质混合物。

20 世纪 60 年代中期，电泳方法的改进版——**SDS-聚丙烯酰胺凝胶电泳**发明，革新了常规的蛋白质分析方法。它采用高度交联的聚丙烯酰胺凝胶作为让蛋白质通过的惰性基质。凝

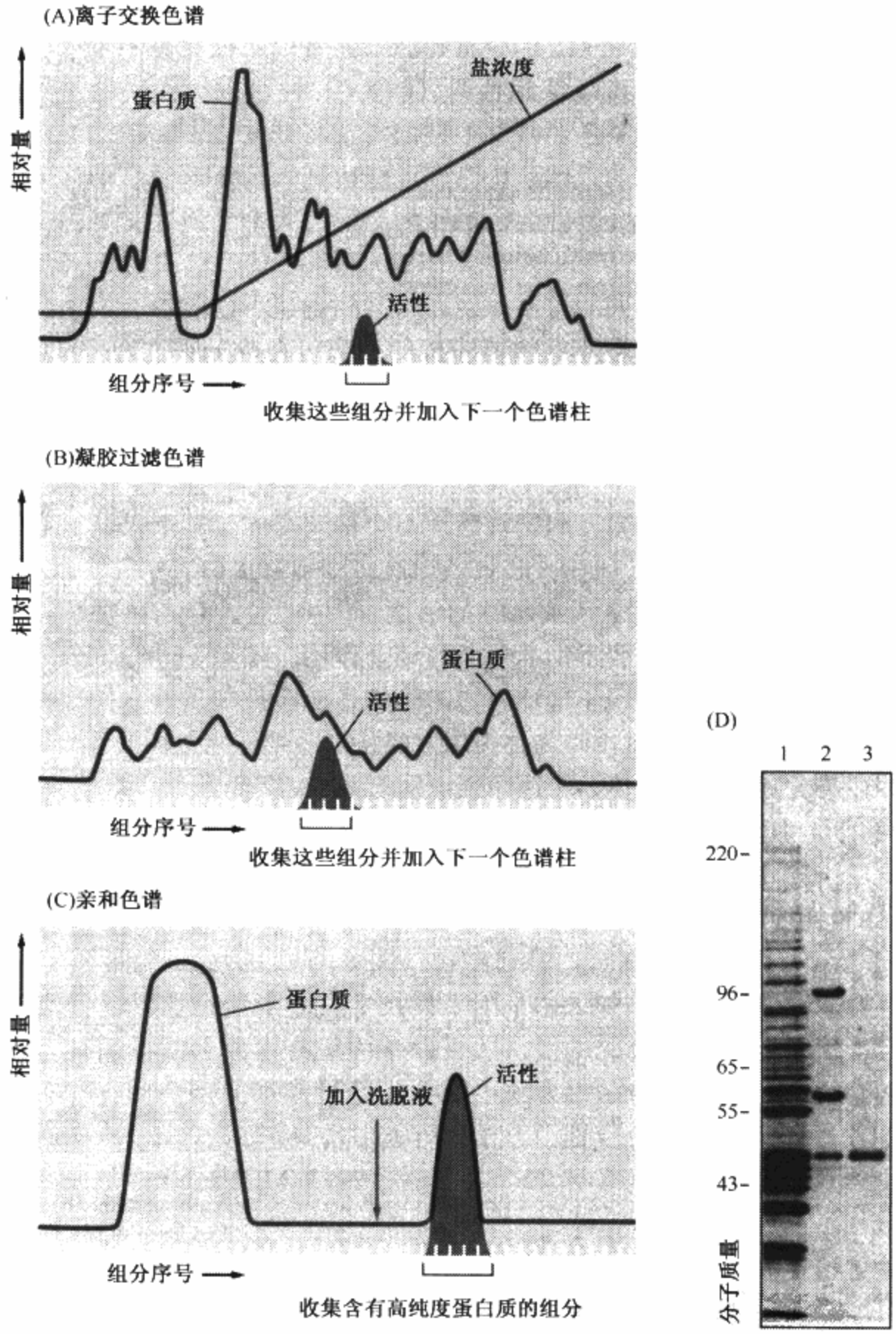


图 8-12　亲和色谱纯化蛋白质

通常要纯化一个蛋白质通过连续的 3 个不同色谱步骤才能得到结果。在这个例子中，细胞匀浆首先通过填充了离子交换树脂的色谱柱进行分离（A）。洗柱后结合蛋白质被从柱上方流下的盐离子浓度不断增加的溶液洗脱。与离子交换树脂亲和性最低的蛋白质是最先洗脱下来的组分。剩下的蛋白质依次洗脱下来——亲和性最高的蛋白质需要最高的盐浓度洗脱。目标蛋白质与几种组分一同洗脱下来，其酶活性被检测到。收集有活性的组分进行下一步的凝胶过滤（B）。不纯的蛋白质洗脱位置再一次通过酶活性确定，收集后在含有酶的固定底物（D）的亲和色谱柱中得到纯化（C）。从裂殖酵母中进行细胞周期蛋白-结合蛋白的亲和纯化（图 8-14）。泳道 1 是细胞的总提取物；泳道 2 是表示蛋白质从含有细胞周期蛋白 B2 的亲和色谱柱中洗脱出来；泳道 3 表示一种主要的蛋白质从细胞周期蛋白 B3 亲和色谱柱中洗脱出来。泳道 2 和泳道 3 中的蛋白质是用盐洗脱的，并用考马斯亮蓝染色［（D）摘自 D. Kellogg et al. *Cell Biol.* 130：675～685，1995.
© The RockefellerUniversity Press］

胶通过单体聚合得到，胶的孔径可以调节，使之足够小适于阻滞目标蛋白质。蛋白质不是存在于水溶液中，而是存在于具有强负电性的去垢剂十二烷基磺酸钠（SDS）溶液中（图 8-13）。因为该去垢剂可以与蛋白质分子的疏水区域相结合，使之展开成伸展的多肽链，所以蛋白质分子从与其他蛋白质或脂类分子的结合中释放出来，完全溶解在去垢剂溶液中。通常要加入 β-巯基乙醇等还原剂破坏二硫键（图 8-13），所以组成多亚基蛋白质分子的所有多肽链可以分别进行分析。

SDS 溶解的蛋白质混合物穿过聚丙烯酰胺凝胶层的时候发生了什么呢？大量带有负电荷的去垢剂分子与蛋白质结合，掩盖掉了蛋白质自身所带的负电荷，使其在外加电场下移向电源的负极。大小相同的蛋白质分子迁移速度相同，因为①它们的天然结构已被打开，所以具有相同的形状；②它们结合了相同数量的 SDS 因而具有相同数量的负电荷。大的蛋白质分子受到更大的电场力，但同时也受到更大的阻力。在自由溶液中，两种效果会相互抵消，但是在充满起着分子筛网作用的聚丙烯酰胺凝胶中，大分子蛋白质受到更强的阻滞。结果混合物中的蛋白质因其相对分子质量的大小顺序分离成一系列彼此分离的蛋白质条带（图 8-14）。大量的蛋白质在含有染料考马斯亮蓝的凝胶中已经显色，微量的蛋白质可用银染或金染处理（灵敏度可达 10ng）。

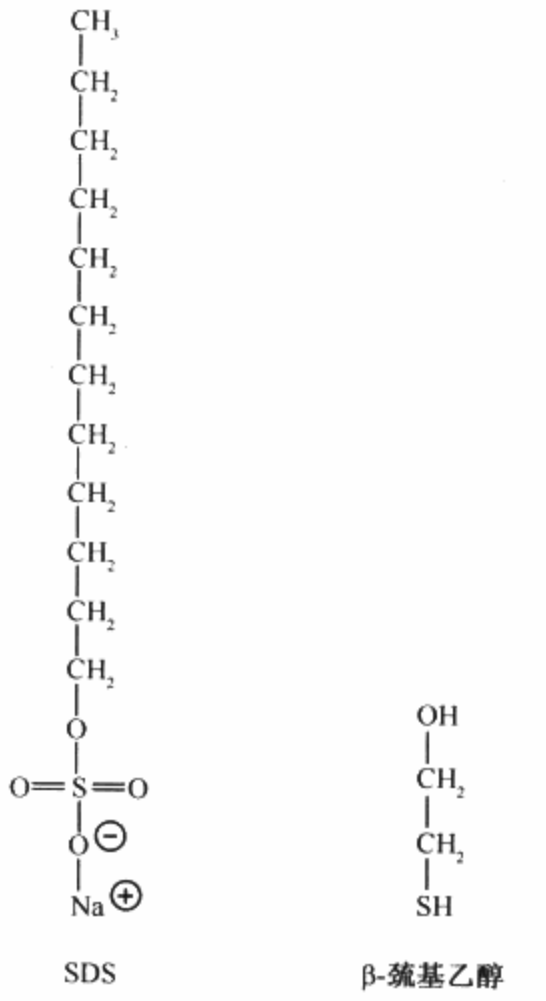


图 8-13　去污剂十二烷基磺酸钠（SDS）和还原剂 β-巯基乙醇

这两种试剂用来在 SDS-聚丙烯酰胺凝胶电泳中稳定蛋白质。SDS 这里展示的是其离子化状态

SDS-聚丙烯酰胺凝胶电泳比以往的蛋白质分析法都有效，主要是因为它可以处理各种类型的蛋白质，包括在水中不溶的蛋白质。膜蛋白、细胞骨架蛋白，及大分子聚合物的组成蛋白质都可以用这种方法。因为这种方法通过分子质量进行分离，所以它还可以提供蛋白质复合体的分子质量和亚基组成信息。图 8-15 中展示了蛋白质提纯过程的一系列的分析步骤中的电泳情况。

超过 1000 种蛋白质可以在一次聚丙烯酰胺双向凝胶电泳中分离

因为靠近的蛋白质条带或峰容易重叠，像 SDS-聚丙烯酰胺凝胶电泳或亲和层析这样的一维分离技术只能用于数量较少的蛋白质（通常少于 50 种）。相反，**双向凝胶电泳**结合了两个不同的分离程序，可以绘出超过 2000 种蛋白质——一个简单细菌中的蛋白质种类总数的二维图谱。

首先，根据蛋白质自身的带电性进行分离。将样品溶解在少量含有非离子（不带电）去污剂、β-巯基乙醇和变性剂尿素的溶液中。溶液可以溶解蛋白质，使之变性，让所有多肽链分开而不改变其电荷。利用在不同的 pH 条件下，蛋白质带电性不同的性质，多肽链在其后的等电聚焦过程中分离。每种蛋白质都有一个特征性的等电点，在等电点的 pH 下蛋白质不带净电荷，因而在电场中不迁移。在等电聚焦中，通过不同缓冲液的混合在聚丙烯酰胺凝胶细管中建立 pH 梯度，将蛋白质电泳分离。每种蛋白质到达等于它等电点的梯度位置之后停止（图 8-16），这是双向电泳的第一向。

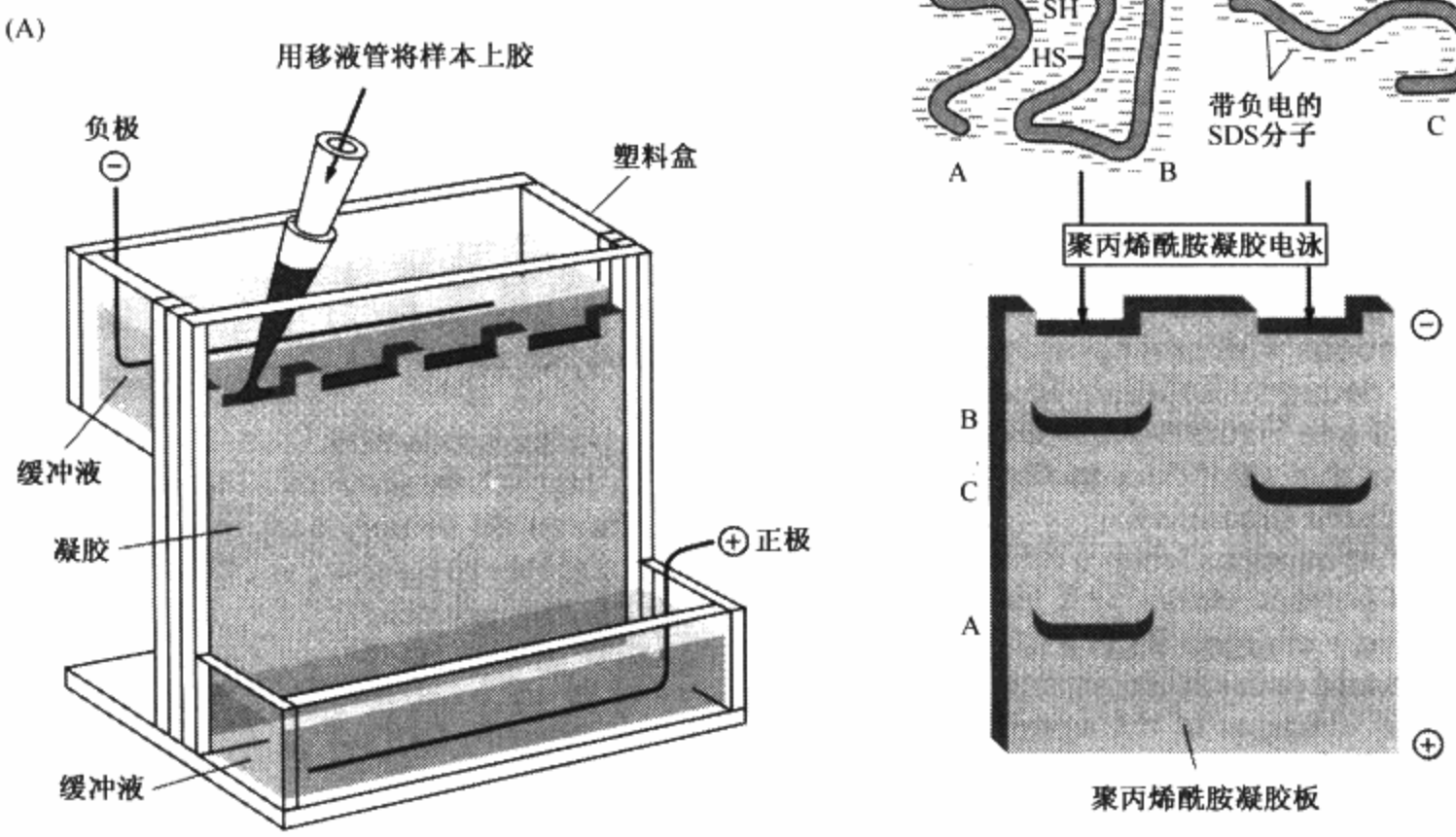


图 8-14　SDS-聚丙烯酰胺凝胶电泳（SDS-PAGE）

（A）电泳装置。（B）单个多肽链与带负电的 SDS 分子形成带负电的 SDS -蛋白质复合物，然后迁移穿过聚丙烯酰胺凝胶的孔。相对分子质量越小多肽链迁移越快，因此这种技术可以测定多肽或蛋白质组成亚基的相对分子质量。但是，如果蛋白质含有太多的糖，它在胶中的迁移就会不规则，SDS-PAGE 得出的表观分子质量就不正确

在第二步中，含有被分离蛋白质的凝胶窄胶在与原先成垂直的方向上再次电泳。这次加入 SDS，使蛋白质依据相对分子质量大小进行分离，如同在单向的 SDS-PAGE 中一样：第一步的窄胶浸过 SDS 之后放在 SDS-聚丙烯酰胺凝胶的边缘，通过它多肽链形成分离的小点。这是聚丙烯酰胺双向凝胶电泳的第二步。此时仍未分离的蛋白质具有相同的大小和等电点，这是很少见的情况。即使是痕量的蛋白质也可以通过各种染色方法检测到，当蛋白质事先进行了放射标记时可以采用放射自显影（图 8-17）。这项技术的分辨率可以达到区分仅仅相差一个带电荷的氨基酸残基的蛋白质。

经过单向或双向电泳，特定的蛋白质可以采用带有放射性同位素、易检测的酶或荧光染料的特异性抗体进行检测。为了方便起见，通常把蛋白质转移到醋酸纤维膜上（用“杂交”的办法），这将在随后的核酸部分进行介绍（图 8-27）。这个蛋白质检测方法叫做 **Western 杂交**（图 8-18）。

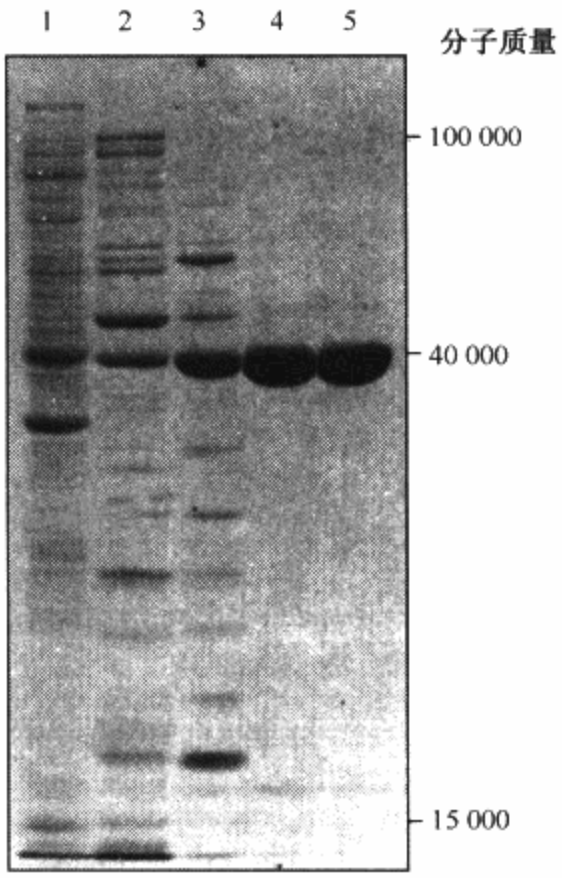


图 8-15　用 SDS-聚丙烯酰胺凝胶电泳法分析蛋白质样品

照片展示的是在提纯一种酶的一系列分析步骤中蛋白质的电泳情况。最左边的泳道（泳道 1）含有初始细胞提取物的复杂蛋白质混合物，接下来的每个泳道含有前一个泳道蛋白质经过亲和层析分离出的一个组分（图 8-12）。等量的蛋白质（10μg）加入到泳道顶部的凝胶中。单个蛋白质通常呈现出锐利的染色带，但含量过高时会变宽。（摘自 T. Formosa and B. M. Alberts, *J. Biol. Chem.* 261: 6107～6118, 1986）

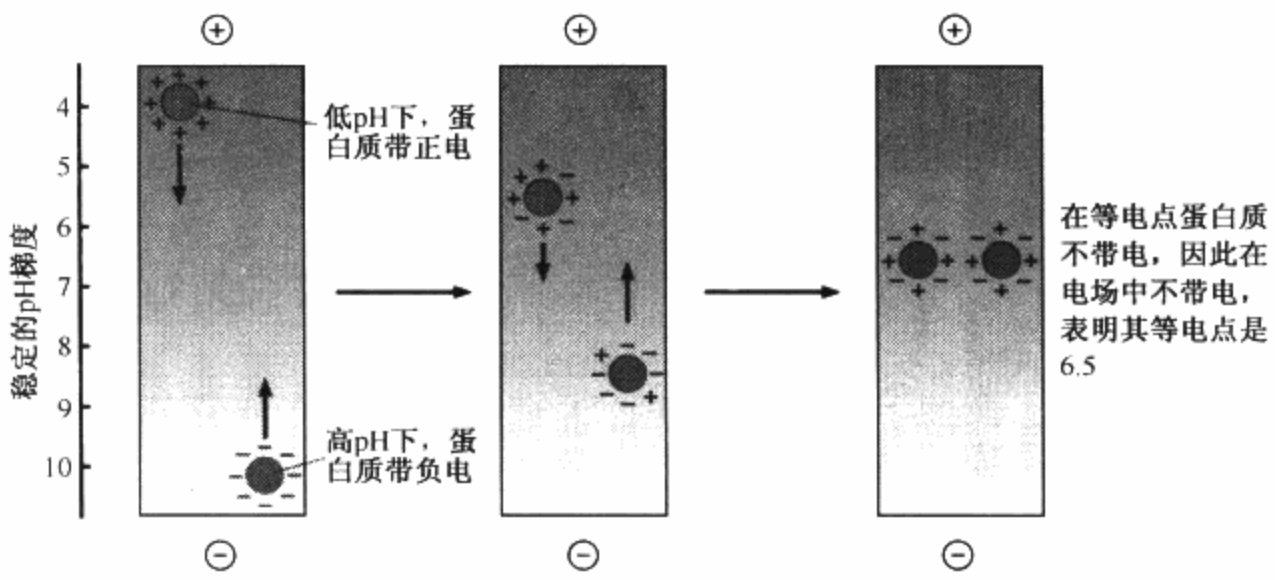


图 8-16　通过等电聚焦分离蛋白质

低 pH（高 H^+ 浓度）下，蛋白质的羧基更多地不带电（—COOH）而含氮碱性基团更多地带正电（如—NH_3^+），使多数蛋白质带上正的净电荷。高 pH（低 H^+ 浓度）下，蛋白质的羧基更多地带负电（—COO^-）而含氮碱性基团更多地不带电（如—NH_2）使多数蛋白质带上负的净电荷。高 pH 下，蛋白质带负电。在等电点正负电荷平衡，蛋白质不带电。因此当有固定 pH 梯度的管在强电场中处于正确的方向时，每一种蛋白质都会在其等电点位置形成一条细带

表 8-5 列出了亲和层析和电泳发展中的一些重要事件。

表 8-5　亲和层析和电泳发展中的重要事件

1833 年	Faraday 提出了电介质溶液中电流通过的基本定律
1850 年	Runge 根据无机化学物质在纸上的吸附性不同进行了分离，这是后来色谱分离技术的前身
1906 年	Tswett 让植物叶子的石油醚提取物通过白垩粉，发明了柱色谱
1933 年	Tiselius 在分离溶液蛋白质时引入了电泳
1942 年	Martin 和 Synge 改进了分离色谱，创造了使用离子交换树脂的纸层析
1946 年	Stein 和 Moore 先采用淀粉进行柱色谱实验，之后又采用离子交换树脂，首次测定了蛋白质的氨基酸序列组成

续表

1955 年	Smithies 使用淀粉凝胶电泳分离蛋白质
1956 年	Ingram 得到了第一个蛋白质指纹，发现镰状红细胞与正常红细胞血红蛋白的差异只有一个氨基酸
1959 年	Raymond 引入了比淀粉凝胶更好的聚丙烯酰胺凝胶电泳分离蛋白质，改进了缓冲系统，使其后的几年中 Ornstein 和 Davis 发明的高分辨率的分离得以实现
1966 年	Maizel 使用十二烷基磺酸钠（SDS）改进蛋白质的聚丙烯酰胺凝胶电泳
1975 年	O'Farrel 把 SDS-聚丙烯酰胺凝胶电泳与基于等电点的电泳结合起来，使用双向凝胶电泳系统分析蛋白质混合物

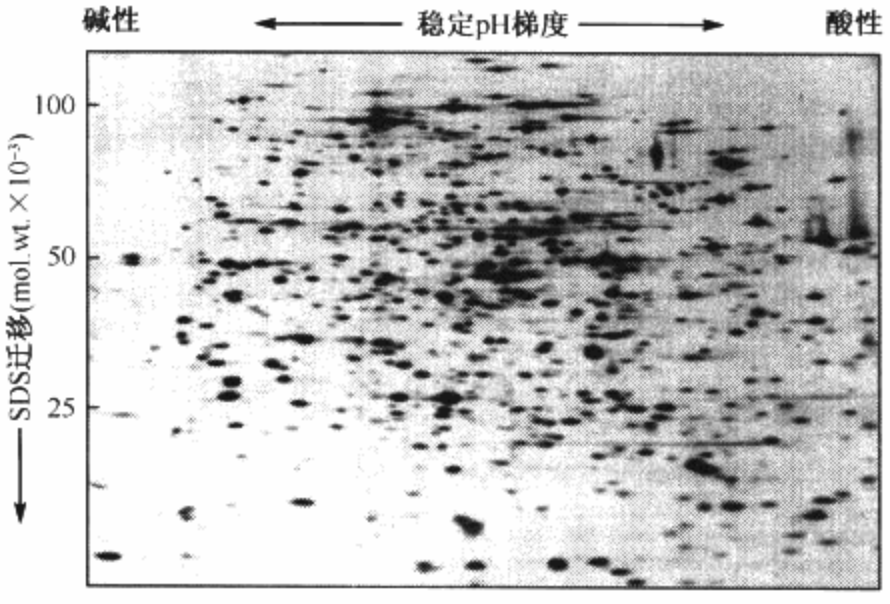


图 8-17　聚丙烯酰胺双向凝胶电泳

大肠杆菌所有的蛋白质都在这块胶上分离了，每一个小点对应一条多肽链。蛋白质先是从左到右根据等电点进行等电聚焦分离。然后在 SDS 存在下进一步依据相对分子质量从上到下进行电泳分离。注意不同的蛋白质含量相差很远。用同位素标记的氨基酸培养细菌后，所有的蛋白质都具有放射性，可以通过放射自显影鉴别（经 Patrck O'Farrell 许可）

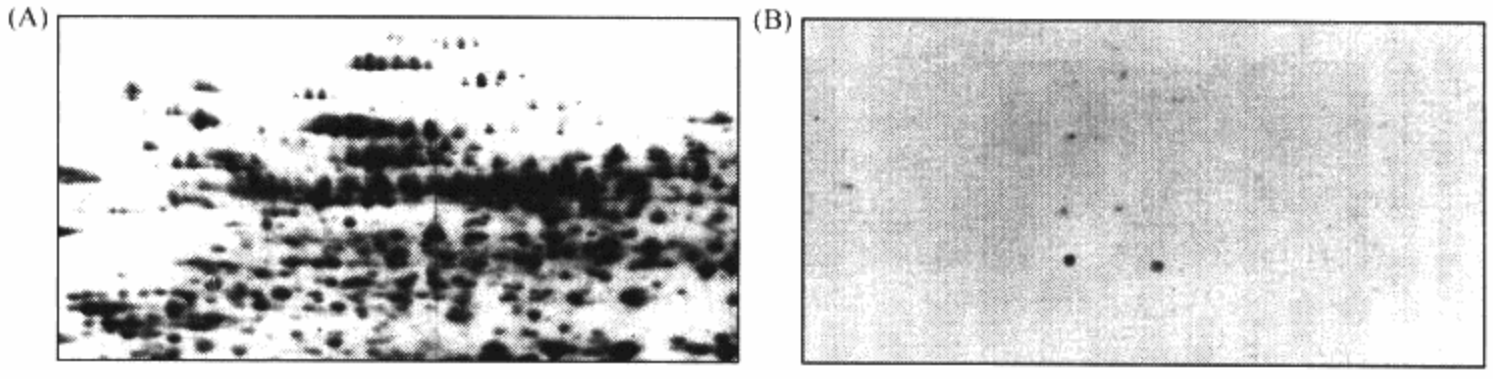


图 8-18　Western 杂交

培养出的分裂期的烟草细胞中的蛋白质，先经双向聚丙烯酰胺凝胶电泳分离（A），分离开的蛋白被转移到硝酸纤维薄膜上（B），然后加入只识别有丝分裂过程磷酸化的苏氨酸残基的抗体。被抗体识别的 10 余个蛋白质经过连有酶的二抗定位。这项技术又称为免疫杂交（摘自 J. A. Traas, A. F. Bevan, J. H. Doonan, J. Cordewener, and P. J. Shaw, *Plant J.* 2: 723～732, 1992. © Blackwell Scientific Publications）

选择性切割蛋白质产生不同的多肽片段组合

尽管蛋白质具有不同的分子质量和等电点，最终准确的判断依赖于它们的氨基酸序列。蛋白质的氨基酸序列可以通过本章后面所介绍的测定核酸序列，并通过遗传密码将其转换成蛋白质序列可以很容易地得到。通过直接分析也可以得到，尽管现在很少直接测定出完整的

蛋白质序列。

有几项更快捷的技术可以提供纯化的蛋白质鉴定中最重要的信息。例如，简单地把蛋白质切割成小片段后可以得到蛋白质分子的一些性质。用水解酶和化学试剂可以在特定的氨基酸残基之间切开蛋白质（表 8-6）。例如，胰蛋白酶在赖氨酸和精氨酸的羧基侧链之间切割，而溴化氰在甲硫氨酸残基旁边切开肽键。因为这些酶和试剂只能选择少数位点，所以当加入到纯化的蛋白质中时，只能产生比较大的肽链。这样的肽链混合物经过色谱或者电泳分离得到的结果可以用于鉴定产生肽链的蛋白质，有时称为蛋白质的“指纹”（图 8-19）。

蛋白质指纹发明于 1956 年，用于对比正常的血红蛋白和镰状细胞贫血症患者的血红蛋白质。单个的肽链的改变被发现，并最终追踪到蛋白质单个氨基酸的变化，首次证明了一个突变可以改变蛋白质中的一个氨基酸。现在这种方法常用于标定翻译后修饰，如磷酸化位点的位置。

从前，切割蛋白质成一组多肽是确定氨基酸序列中重要的一步。人们通过重复一系列的化学反应逐次从每条多肽的 N 端切去氨基酸得到最终的序列。在每一个周期之后，切掉的氨基酸可以用色谱的方法鉴定。目前很多生物的基因组序列已经得到，可以用质谱的方法鉴定蛋白质，并与相应的基因匹配从而得到它的氨基酸序列。我们下面就讨论。

表 8-6　切割蛋白质肽键常用的一些试剂

	氨基酸 1	氨基酸 2
酶		
胰蛋白酶	Lys 或 Arg	任何
糜蛋白酶	Phe、Trp 或 Tyr	任何
V8 蛋白酶	Glu	任何
化学试剂		
溴化氰	Met	任何
2-硝基-5-氰基苯甲酸	任何	Cys

注：切开的键两侧的氨基酸特异性都已指出。氨基酸 2 连在氨基酸 1 的 C 端。

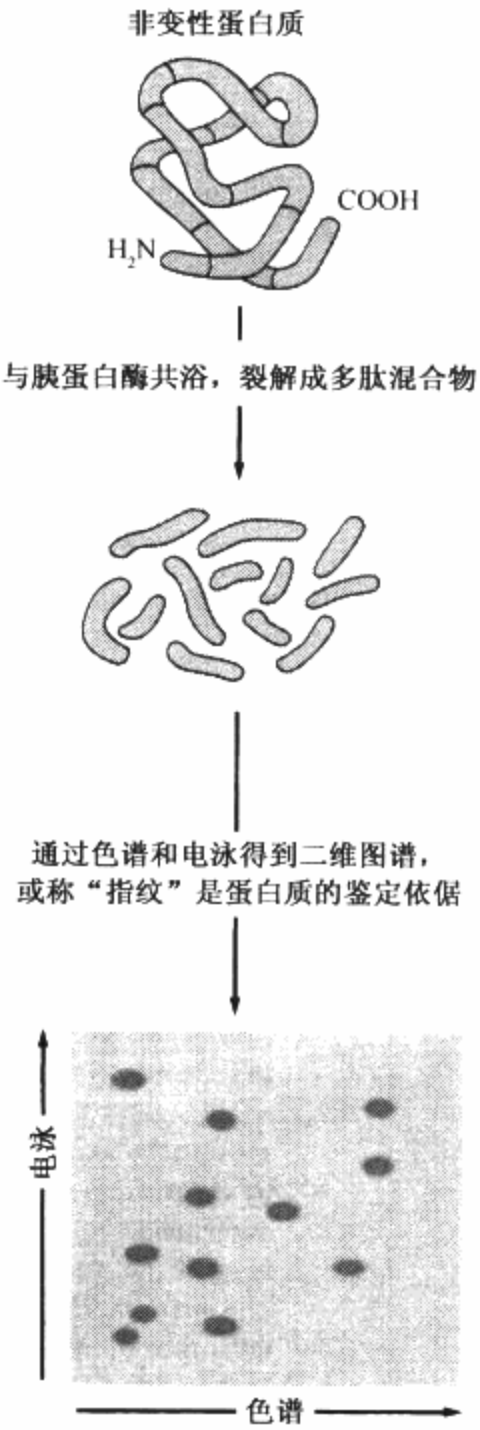


图 8-19　一个蛋白肽谱或称指纹的产生

在这里，蛋白质经胰蛋白酶水解生成多肽混合物，通过电泳和分配色谱在二维尺度上分离。后一项技术根据的是多肽在水中不同的溶解性，与它们的溶剂相比，它们更倾向于与固相基质结合。得到的图谱上的小点来自于被分析蛋白质的特异性水解物。这种方法也可以用于检测蛋白质的翻译后修饰

质谱可以用来进行肽段测序和蛋白质鉴定

质谱可以用来确定完整蛋白质或者经化学切割或酶切所得的多肽的精确分子质量。这个信息可以用于在基因组数据库进行搜索，库中列出了每种蛋白质及推测出的它们所能产生的肽链的分子质量图［图 8-20（A）］。通常知道了一种给定蛋白质的几个肽段分子质量就可以精确定出一个特异的可读框。因此，质谱的方法在蛋白质组学——在生物的基因组中定位并

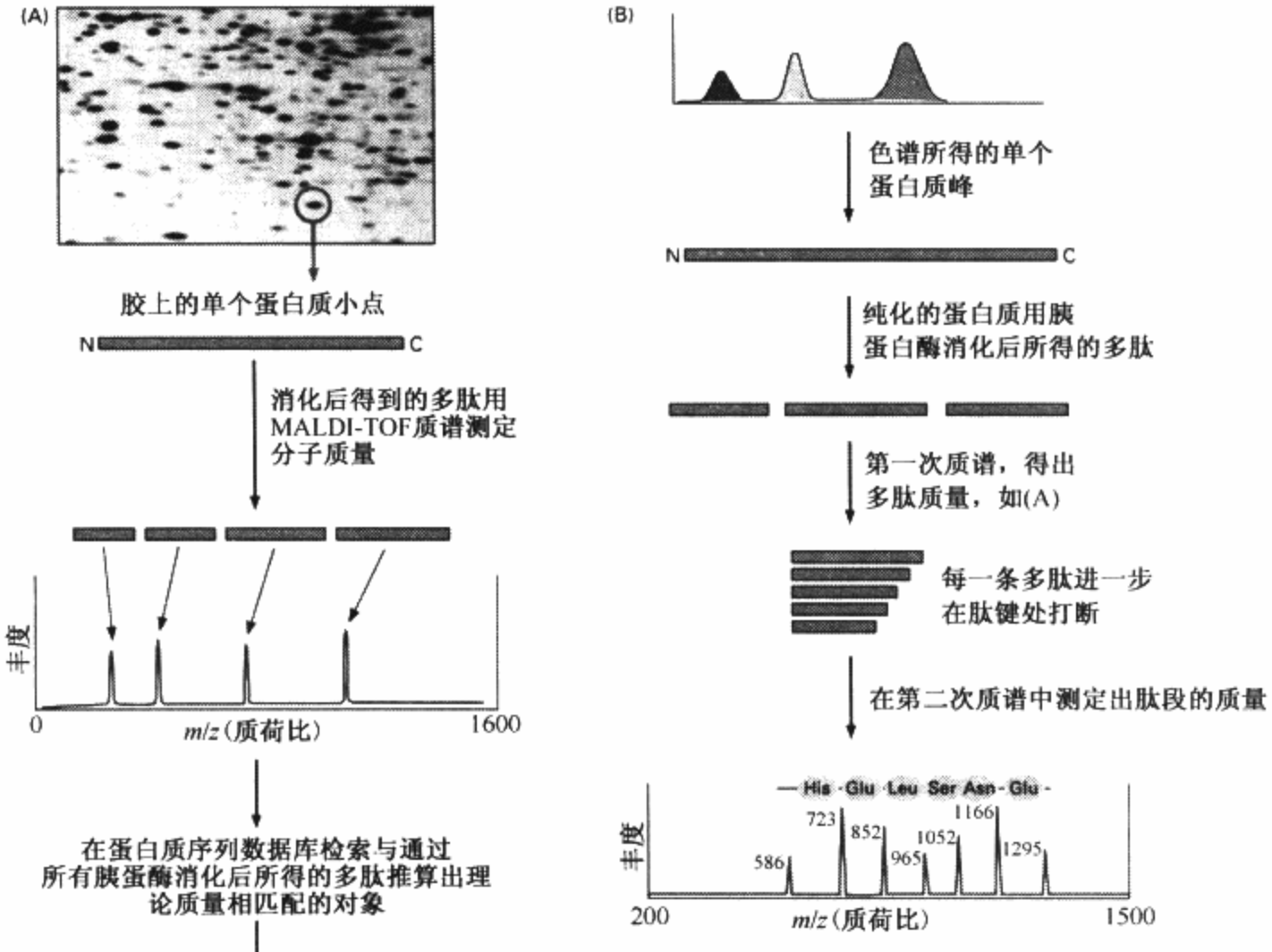


图 8-20　用质谱法鉴定蛋白质和多肽测序

（A）利用质谱，通过测定蛋白质产生的多肽的精确质量可以鉴定蛋白质。利用该信息可以在基因组数据库中检索相匹配的基因。在这个例子中，所研究的蛋白质通过双向凝胶电泳分离得到，并用胰蛋白酶消化。产生的肽段通过质谱测得分子质量。在序列数据库中检索出典型酶酶切参数相匹配的蛋白质编码基因。（B）质谱也可用于直接测定多肽的氨基酸序列。在这个例子中，通过色谱分离得到形成大分子复合体的蛋白质，单个蛋白质用胰蛋白酶消化。酶切所得的肽段用（A）中的质谱方法测定分子质量。为得到精确的氨基酸序列，主要在肽键处进一步打碎每条多肽。处理产生一系列依次相差一个氨基酸的多肽片段。把这些片段加入第二台质谱仪测定分子质量。重复这一过程，可以得到原蛋白质的部分氨基酸序列（显微照片引自 Patrick O'Farrell）

描述所有蛋白质，包括它们翻译后修饰的浩繁工作中非常重要。

质谱需要的样品量少，灵敏度极高。得到的分子质量非常精确，误差通常不超过 1/100 万。最常用的质谱法叫做基质辅助激光解吸离子化-飞行时间质谱（MALDI-TOF MS）中。在这种方法中，多肽与一种有机酸混合后涂在金属或者陶瓷载片上晾干。用激光轰击样品，使蛋白质以电离的气体形式被轰击出去，每分子带有一个或更多的正电荷。此后，离子化的多肽在电场中加速，飞向检测器。它们到达检测器所需的时间由它们的质量和所带的电荷决定：相对分子质量大则移动慢，电荷多则移动快。带有单电荷的分子的质量可由此精确测定。MALDI-TOF 可测定分子质量高达 200 000Da，相当于大约 2000 个氨基酸长度的蛋白质。

质谱也可以用于确定单个肽段的氨基酸序列。这种方法在所研究的生物的基因组序列还不完全知道的情况下尤其有用，这个过程中所确定的部分序列可以用于鉴定和克隆该基因。多肽测序对于被修饰的蛋白质，如糖基化的、磷酸化的或者甲基化的蛋白质也非常有用。在这种情况下，可以精确确定被修饰位点所在的氨基酸。

为了得到多肽的序列信息，先后需要两台质谱仪。一台用来分离所研究的蛋白质消化后产生的多肽，每次分出一个多肽。这个多肽进一步用高能气体原子撞碎。这种破碎方法倾向于在肽键上打断多肽，产生一系列依次相差一个氨基酸的梯度肽段。第二台质谱仪分离这些肽段并测出它们的分子质量。从多肽片段之间的分子质量差异可以推导出氨基酸的序列［图8-20（B)］。

为了进一步了解一种蛋白质的功能，人们需要得到大量的蛋白质以进行分析。通常采用下面介绍的DNA重组技术可以有效达到这个目的。

小结

通过破碎、分离和超速离心其内容物的方法可以对细胞群进行生物化学分析。进一步分离使得研究复杂细胞过程的细节所需的有功能活性的非细胞体系的产生成为可能。蛋白质合成、DNA复制、RNA剪切、细胞周期、有丝分裂以及各种胞内运输过程都可以通过这种途径进行研究。只用少量的蛋白质就可以通过SDS-聚丙烯酰胺凝胶电泳确定其分子质量和亚基组成。在双向凝胶电泳中，依据蛋白质的等电点在第一维分离蛋白质，然后再在第二维上进行SDS-聚丙烯酰胺凝胶电泳。即使不溶于水的蛋白质也可以用这些电泳方法分离。

细胞可溶性提取物中的大量蛋白质可以通过柱色谱纯化。根据色谱柱介质的类型，具有生物化学活性的蛋白质可以通过分子质量、疏水性、带电性以及与其他分子的亲和性差异分离开。在常规的纯化过程中，样品要依次通过多个不同的色谱柱——一个柱中富集的组分加入到下一个柱中。一旦得到蛋白质的纯品，就可以用于生物化学活性的细致研究。应用质谱可以快速测定蛋白质和它所产生的多肽的分子质量。利用这些信息可以在基因组数据库中检索蛋白质，从而根据其基因的序列推断出氨基酸序列。

分离、克隆DNA和DNA测序

直到20世纪70年代DNA还是最难进行生物化学分析的细胞物质。构成生命体遗传物质的超长的、化学性质单一的DNA链只能通过间接的方式进行研究，如蛋白质和RNA的测序或遗传分析。今天情况完全改变了。DNA从最难分析的细胞大分子变成了最易分析的。现在可以从一个基因组中分离出一个特定区域，产生无数的拷贝，并在一夜之间测定它的核酸序列。在人类基因组计划中，众多的工作人员使用自动化的仪器可以不停地每秒产生1000个核酸。使用相关技术，可以根据需要对分离的基因进行改造，并转回到动物或植物的种系中，成为生物体基因组有功能的、可遗传的组成部分。

遗传工程的技术突破（在试管或生物体中精确操纵DNA的能力）以此前难以想像的方式促进了对细胞及其大分子的研究，对细胞生物学的各个方面都产生了巨大的影响。这些技术促成了新基因和蛋白质类群的发现，并表明很多蛋白质在进化中的保守性远比人们预测得要高。它们为鉴定蛋白质和蛋白质中的单个结构域的功能提供了新工具，揭示出它们之间许多未预料到的关系。这些技术使大量地产生任意蛋白质成为现实，为高效生产蛋白质激素和疫苗提供了出路。最后，通过切割基因的调控区，这些技术突破为生物学家提供了搞清复杂真核基因表达调控网络的工具。

DNA重组技术由各种技术组成，有些是新的，有些是从微生物遗传学等其他领域借鉴而来的（表8-7)。其中最核心的技术如下：

表 8-7　DNA 重组技术和转基因技术发展中的重大进展

1869 年	Miescher 首次从附近医院沾有脓液的绷带上收集白细胞，分离出 DNA
1944 年	Avery 提供证据表明在细菌转化过程中，携带遗传信息的是 DNA 而不是蛋白质
1953 年	Watson 和 Crick 基于 Franklin 与 Wilkins 的 X 射线衍射结果提出了 DNA 双螺旋结构模型的假说
1955 年	Kornberg 发现了今天用来制造标记的 DNA 探针的 DNA 聚合酶
1961 年	Marmur 和 Doty 发现了 DNA 复性，是核酸杂交的特异性和可行性的基础
1962 年	Arber 提供了 DNA 限制性内切核酸酶存在的第一个证据，随后 Nathans 和 H. Smith 对其进行了纯化并在 DNA 序列的鉴定中使用
1966 年	Nirenberg、Ochoa 和 Khorana 解出遗传密码
1967 年	Gellert 发现了用于把 DNA 片段连在一起的 DNA 连接酶
1972～1973 年	Boyer、Cohen、Berg 的实验室及其在斯坦福大学和加州大学旧金山分校的合作者发明了 DNA 克隆技术
1975 年	Southern 发明了用凝胶转移杂交法检测特定 DNA 片段的方法
1975～1977 年	Sanger 和 Barrell、Maxam 和 Gilbert 发明了 DNA 快速测序法
1981～1982 年	Palmiter 和 Brinster 培养出转基因小鼠，Spradling 和 Rubin 培养出转基因果蝇
1982 年	美国国立卫生院的基因序列数据库 GenBank 在 Los Alamos 国家实验室建成
1985 年	Mullis 及其合作者发明了聚合酶链反应
1987 年	Capecchi 和 Smithies 创造了在小鼠胚胎干细胞中进行标记基因替换的方法
1989 年	Fields 和 Song 发明了用于鉴定蛋白质相互作用的酵母双杂交系统
1989 年	Olson 及其合作者提出了序列标记位点，是用来制作人类染色体物理图谱的 DNA 的独特的延伸序列
1990 年	Lipman 及其合作者建成了 BLAST 算法，用来对 DNA 和蛋白质序列的同源性进行检索
1990 年	Simon 及其合作者研究如何高效利用细菌人工染色体（BAC）携带克隆的大片段人类 DNA 用于测序
1991 年	Hood 和 Hunkapillar 发明了新的 DNA 自动测序技术
1995 年	Venter 及其合作者完成了第一个完整基因组流感细菌（*Haemophilus influenzae*）的测序
1996 年	Goffeau 与一个国际协会的研究者一起宣布完成了第一个真核生物裂殖酵母（*Saccharomyces cerevisiae*）全基因组的测序
1996～1997 年	Lockhart 及其合作者及 Brown 和 DeRisi 发明了 DNA 微阵列，可同时追踪上千基因
1998 年	Sulston 和 Waterston 及其合作者提供了第一个真核生物秀丽新小杆线虫（*Caenorhabditis elegans*）的全序列
2001 年	国际研究者宣布人类基因组草图完成

1. 用限制性内切核酸酶在特定位点切割 DNA，大大方便了单个基因的分离和操纵。

2. 通过使用载体或通过聚合酶链反应克隆 DNA，使得一个 DNA 分子可以被拷贝扩增到数十亿个相同的分子。

3. 基于与互补的核酸序列结合能力的核酸杂交，可以以很高的准确性和灵敏度寻找特异的 DNA 或者 RNA。

4. 纯化的 DNA 片段的快速测序，可以用来鉴定基因和推断其编码的蛋白质的序列。

5. 用核酸微阵列同时进行数万个杂交反应，追踪细胞中每个基因的表达水平。

在这一章中，我们将讨论以上使细胞生物学产生革新的每一项基本技术。

可以用限制性内切核酸酶将大分子 DNA 切成片段

一个基因不像一个蛋白质那样分散地存在于细胞中，而是作为一个长 DNA 分子的一个区域存在。尽管可以用机械力把 DNA 分子打成随机片段，但是在哺乳动物基因组中 10 万个

以上这样的片段中才含有一个基因，以它们的平均大小是难以分辨的。怎样才能分辨这样一个基因呢？因为 DNA 分子大致都含有相等含量的 4 种核苷酸，所以不能像蛋白质那样根据电荷和结合性质进行分离。而且，即使能够进行纯化，为了进行进一步实验还需要大量的 DNA。

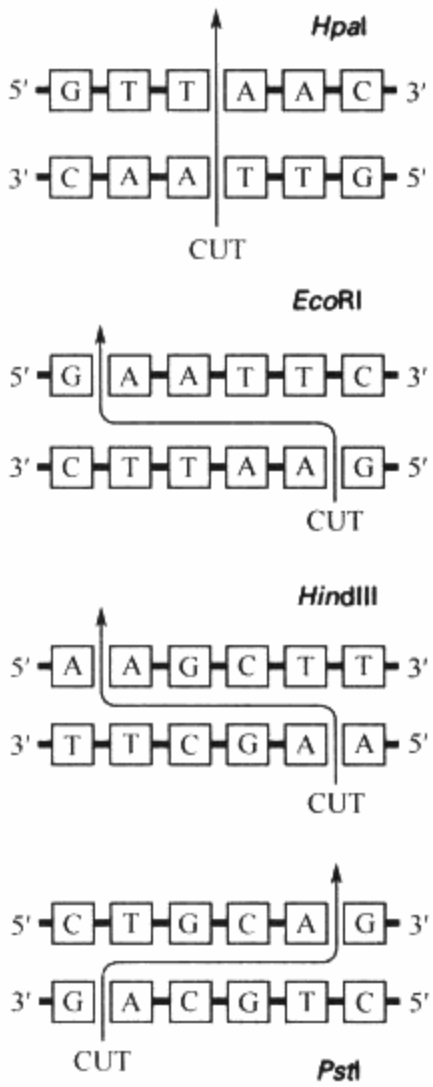


图 8-21　4 种常用的限制性内切核酸酶识别的核酸苷序列

如图中所示，这些序列通常是 6bp 长的回文序列（即绕识别位点中心旋转 180°后序列不变）。酶在 DNA 两条链的识别序列处或其附近切割。有些酶，如 *Hpa* Ⅰ，切割后产生平末端；另一些酶，如 *Eco*R Ⅰ、*Hind* Ⅲ和 *Pst* Ⅰ，切割产生不齐的黏性末端。不同的限制性内切核酸酶来源于不同的细菌：*Hpa* Ⅰ来源于 *Hemophilus parainfluenzae*，*Eco*R Ⅰ来源于 *Escherichia coli*，*Hind* Ⅲ来源于 *Hemophilus influenzae*，*Pst* Ⅰ来源于 *Providencia stuartii*

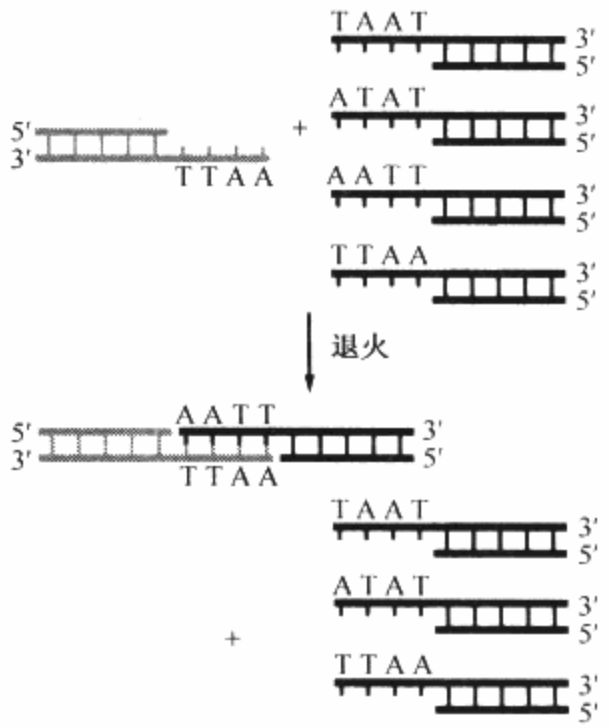


图 8-22　限制性内切核酸酶产生的 DNA 片段可以容易地连接在一起

如图所示，具有相同黏性末端的片段可以通过碱基对互补容易地结合在一起。在这个例子中，结合的两个 DNA 片段都是 *Eco*R Ⅰ切割产生的，而另外的 3 个片段都由其他的限制酶产生，具有不同的黏性末端（图 8-21）。平末端的片段，如 *Hpa* Ⅰ产生的片段（图 8-21），也可以接在一起，但是更为困难

限制性内切核酸酶的发现使得所有的问题都开始得到解决。这些可从细菌中纯化的蛋白质，在 DNA 双螺旋的特定位点切割，产生大小确定的双链片段，其切割位点由核苷酸序列决定。不同的限制性内切核酸酶具有不同的序列特异性，寻找能切割出包含特定基因片段的限制酶相对比较容易。该 DNA 片段的大小可以作为从混合物中分离该基因的一个依据。

不同种类的细菌产生不同的限制酶，可以通过降解外源病毒 DNA 保护它们免受病毒侵害。每种限制酶识别 DNA 中 4～8bp 的一个特异序列。在细菌自己的基因组出现的这些序列则在 A 或 C 残基处被甲基化，因而受到保护，免于被切割。外源的 DNA 通常不能被甲基化，

所以被限制酶切割。很多限制性内切核酸酶都从细菌中被纯化出来，数百种大多识别不同核酸序列的限制酶已经商品化。

有些限制性内切核酸酶产生不整齐的切口，每一片段的末端留有单链的末端（图 8-21）。这样的末端叫做黏性末端，每个这样的末端可以与用同样的酶切出的任一片段末端互补（图 8-22）。黏性末端使得两个用同样的限制酶（或产生同样黏性末端的不同限制酶）切出的两个 DNA 片段可以结合在一起。切割两个或两个以上 DNA 片段产生的 DNA 分子叫做**重组 DNA** 分子，它们使很多新的细胞生物学研究得以实现。

凝胶电泳分离不同大小的 DNA 分子

采用在蛋白质分析中十分有用的凝胶电泳方法同样可以鉴定 DNA 分子的长度和纯度。甚至过程比蛋白质更简单：因为核酸分子中的每个核苷酸本身就带有一个负电荷，所以不用再加入使蛋白质均一地向电极正极移动的、带负电的去垢剂 SDS。对于小于 500bp 的 DNA 片段，采用特别设计的聚丙烯酰胺凝胶电泳可以分辨出仅相差一个核苷酸的不同长度的分子［图 8-23(A)］。聚丙烯酰胺凝胶的孔径太小，不允许相对分子质量太大的 DNA 分子通过，所以要根据长度分离这样的 DNA，需要采用更多孔的可溶性琼脂糖的（一种从水草中分离出的多糖）凝胶［图 8-23(B)］。这些 DNA 分离方法广泛地应用在分析和制备中。

一种特殊的琼脂糖凝胶电泳——脉冲凝胶电泳，可以用于分离超长的 DNA 分子。常规凝胶电泳不能分离这些分子，因为稳定的电场使得它们伸直，以尾朝前的方式在凝胶中以不依赖于长度的速率做蛇形运动。然而在脉冲凝胶电泳中，电场的方向是周期性改变的，迫使分子不断重新定向，之后才能在胶中做蛇形运动。较大的大分子重新定向的时间更长，所以比较小的小分子移动得更慢。因此，细菌或酵母的整条染色体都可以在脉冲凝胶电泳中分成分离的条带并通过相对分子质量进行分类和鉴定［图 8-23(C)］。尽管通常的哺乳动物染色体长达 10^8bp，即使用这种办法也不能分析，但是经过寡酶切位点（每 10 000bp 或以上一个酶切位点）的限制酶的切割的染色体大片段可以用这种方法分离和鉴定。

琼脂糖或聚丙烯酰胺凝胶中的 DNA 条带只有标记或染色后才能看到。一种灵敏的染色法是加入溴化乙锭，它与 DNA 结合后可在紫外光下发荧光［图 8-23(B)、(C)］。另一种更为灵敏的方法是在电泳前在到 DNA 中掺入放射性同位素，常常把 ^{32}P 掺入到 DNA 的磷中，使其发出 β 射线，可以很容易地进行放射性自显影检测［图 8-23(A)］。

纯化的 DNA 分子可在体外用同位素或化学标签进行特异标记

有两项技术广泛用于标记分离的 DNA 分子。在第一项技术中，DNA 聚合酶在有放射性的（通常用 ^{32}P 标记）或进行了化学标记的核苷酸存在下复制 DNA［图 8-24(A)］。这样，产生的含有许多标记了的核苷酸的“DNA 探针”可与核酸进行杂交反应（在下面讨论）。第二项技术采用噬菌体的多聚核苷酸激酶从 ATP 上转移一个 ^{32}P 标记的磷酸基团到每个 DNA 链的 5′端［图 8-24(B)］。因为只有一个 ^{32}P 原子被激酶转移到每个 DNA 分子上，所以用这种方法标记的 DNA 分子的放射性不够作为 DNA 探针使用。但是因为它们只在一端标记，所以对于包括我们将简要介绍的 DNA 足迹实验在内的其他应用非常有用。

今天，放射性标记方法正在被化学和荧光标记法所取代。使用特殊修饰的核苷酸前体可以制造这样的非放射性 DNA 分子［图 8-24(C)］。这样制造的 DNA 分子可以通过杂交与它的互补 DNA 序列结合（我们将在下一节讨论），之后可用特异性识别其被修饰的侧链的抗体（或其他配体）进行检测（图 8-28）。

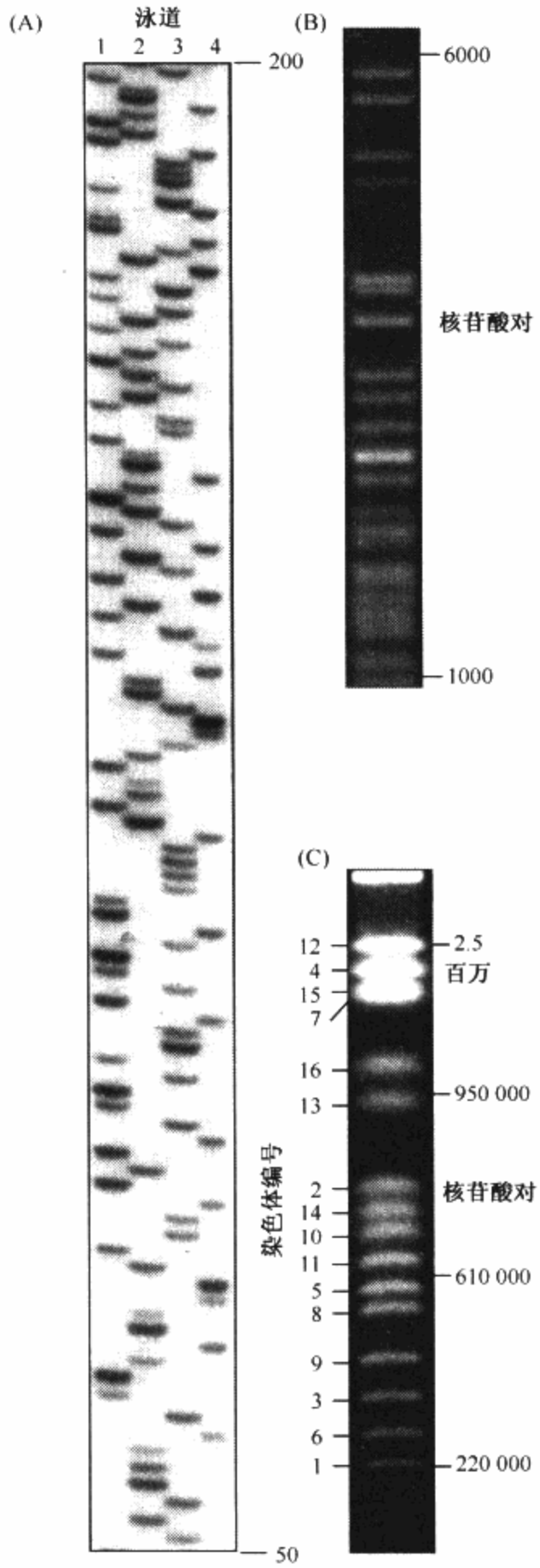


图 8-23　使用凝胶电泳技术根据大小分离 DNA

在展示的 3 个例子中，电泳的方向是从上到下的，因此最大的，即移动最缓慢的 DNA 分子位于凝胶的最顶端。（A）小孔径的聚丙烯酰胺凝胶用来分离单链 DNA。大小为 10～500nt 的 DNA 分子可以因一个碱基的差别而分离开。示例中的 4 条泳道是 DNA 测序过程中合成的 4 个体系。要测序的 DNA 可以从一个固定的起始位点开始人工合成，在不同的位置终止，产生一系列不同长度的部分复制链（图 8-36 阐释了这些部分复制链的合成过程）。泳道 1 展示的是终止于 G 的部分复制链，泳道 2 展示的是终止于 A 的部分复制链，泳道 3 展示的是终止于 T 的部分复制链，泳道 4 展示的是终止于 C 的部分复制链。因为反应中的 DNA 分子进行了放射性标记，所以它们的位置可以通过放射自显影判断。（B）中孔径的琼脂糖凝胶电泳用来分离双链 DNA 分子。这种方法对于长度为 300～10 000bp大小的 DNA 最有效。这些 DNA 分子是用限制性内切核酸酶切割细菌病毒基因组后得到的片段，可以用溴化乙锭染色后经荧光显色进行检测。（C）脉冲琼脂糖凝胶电泳技术用于分离酵母（*Saccharomyces cerevisiae*）的 16 条染色体，大小为 220 000～2500 000bp（图 4-13），用（B）中的方法染色。不大于 10^7bp 的 DNA 可以采用这种方法［（A）经 Lauffer 和 Peter Walter 许可；（B）经 Ken Kreuzer 许可；（C）摘自 D. Vollrach and R. W. Davis, *Nucleic Acids Res.* 15：7865～7876，1987. © Oxford University Press］

核酸杂交反应是检测特异性核苷酸序列的灵敏手段

当 DNA 的水溶液被加热到 100℃或者暴露在很高的 pH（pH≥13）下时，在通常情况下互补形成双螺旋的两条链上的碱基对就会被打开，双螺旋迅速分离成两条单链。这个过程叫做 DNA 的变性，在很长时间里这一过程被认为是不可逆的。但是在 1961 年，人们发现在 65℃长时间保温可使互补的 DNA 单链很好地重新形成双螺旋，这个过程叫做**杂交**（也叫 DNA 的复性）。相似地，含有互补的核苷酸序列的任意两条核酸单链（DNA/DNA、RNA/

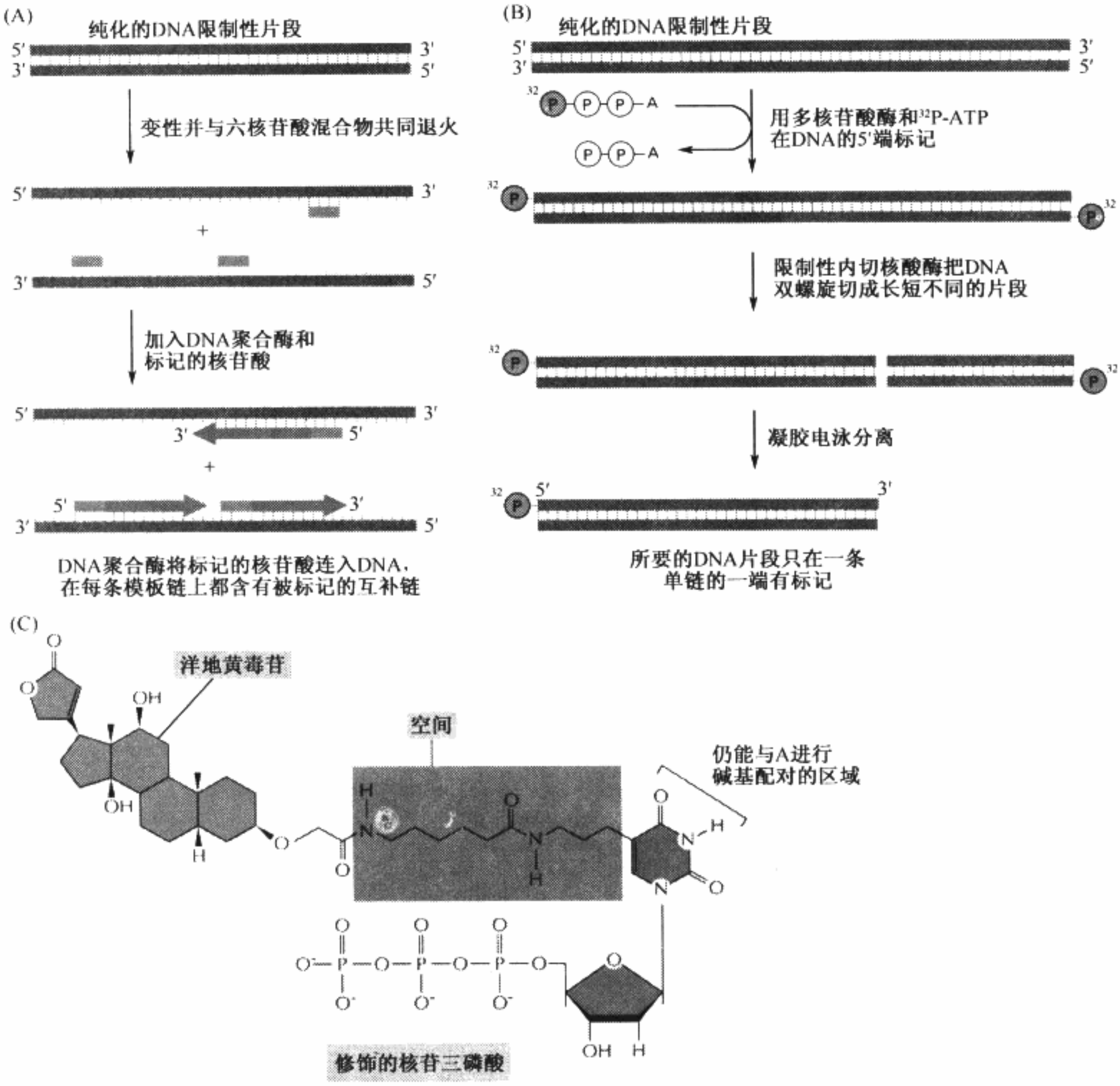


图 8-24　体外标记 DNA 分子的方法

（A）纯化的 DNA 聚合酶标记一个 DNA 分子中的每个核苷酸，从而产生高放射性的 DNA 探针。（B）多核苷酸激酶仅在 DNA 链的 5′端进行标记，之后用限制性内切核酸酶切割可得到含有 5′端标记的单链 DNA 分子。（C）和（A）中的方法也用于产生非放射性的、携有特异性化学标记的 DNA 分子，可以用适当的抗体检测。图示中经修饰的核苷酸被 DNA 聚合酶连接进 DNA 中，使 DNA 分子作为探针能够容易被检测到。图示中核苷三磷酸上的碱基是胸腺嘧啶的类似物，胸腺嘧啶的甲基基团被替换成一个连接着植物激素洋地黄毒苷的空间臂。用特异的连有如荧光染料等可视基团的抗体可以检测洋地黄毒苷。其他的化学标记，如生物素也可以连到核苷酸上，以大致相同的方法使用

RNA 或 RNA/DNA）都可以发生杂交。这种特异的杂交反应被广泛地用于检测和分析 RNA 和 DNA 分子中的特定序列。

用于检测互补序列的单链 DNA 分子通常叫做**探针**。这些带有用于检测的放射性或者化学标记的分子，可以是 15 到数千核苷酸的任意长度。用探针进行的杂交反应灵敏度可达到检测出每个细胞一个分子浓度的互补序列。因此可以判断出特定 DNA 样品中的 DNA 片段拷贝数。同样的方法可以用于寻找相关但不完全相同的基因。例如，以已知的基因为探针，可以在所有基因组还没测序的生物体中找到感兴趣的基因（图 8-25）。

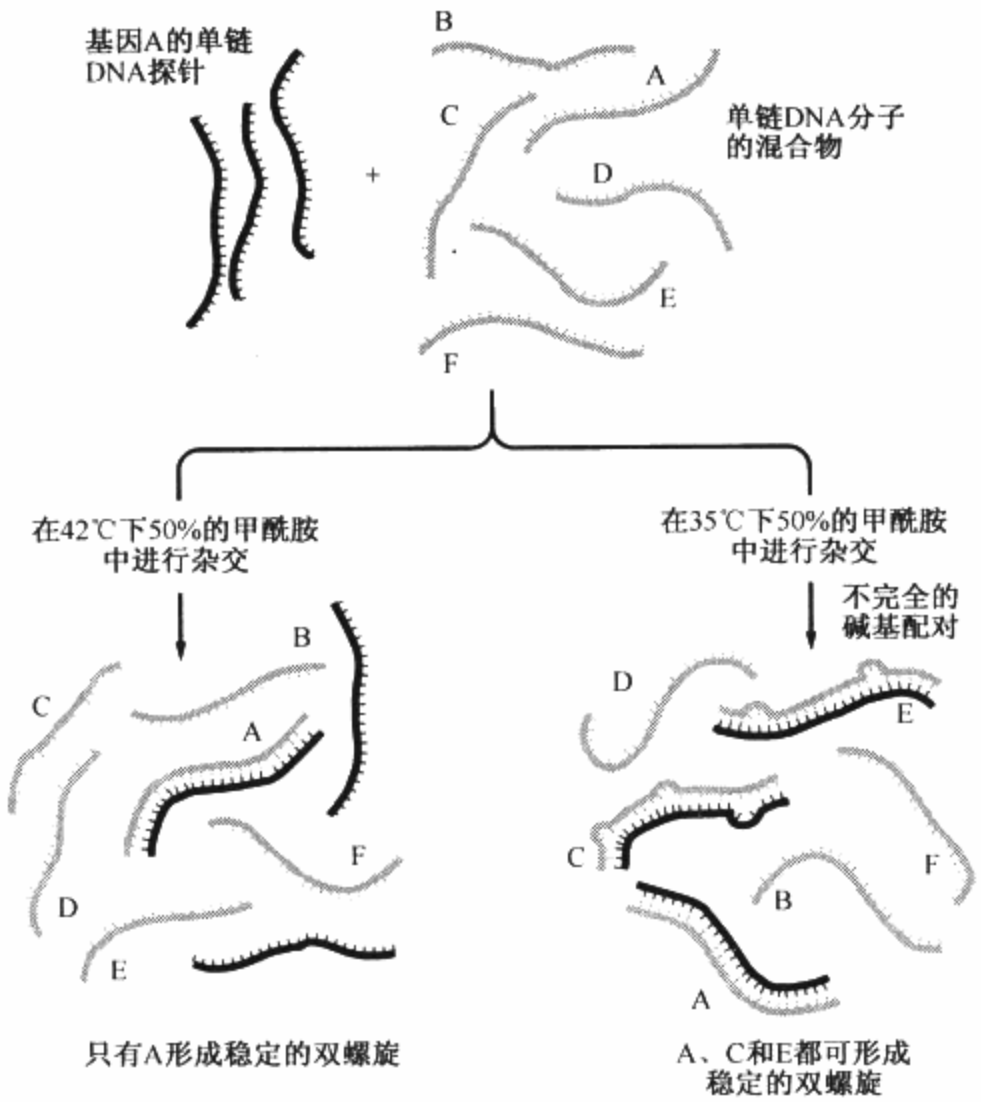


图 8-25　不同的杂交条件下可以进行 DNA 的不精确配对

当人们只需要与探针完全匹配的 DNA 时，杂交温度控制在完好的 DNA 双螺旋在溶液中变性（即熔化）温度之下几度，这样不完全匹配的双螺旋不能稳定存在。当要寻找的是与探针序列相关而并非完全匹配的 DNA 时，杂交温度设置得更低，使得不完全匹配的双螺旋也可以形成。只有较低的杂交温度可以用来搜索与基因 A 有关但不一样的基因（例子中的 D 和 E）（图 10-18）

DNA 探针也可以用于与 RNA 进行杂交，来确定一个细胞中是否表达一个特定的基因。在这种情况下，含有基因序列的 DNA 探针与从细胞中提纯的 RNA 进行杂交，来检测 RNA 中是否含有与 DNA 探针匹配的分子，如果有的话，含量是多少。在更精确的反应中，在杂交反应之后，还要用特异的核酸酶处理 DNA 探针，以判断与细胞中 RNA 配对的 DNA 探针的精确区域。因此，人们可以判断出 RNA 转录的起始和终止位点，也可以找出基因中内含子和外显子的精确边界（图 8-26）。

现在，内含子/外显子的边界通常通过表征了 mRNA 序列的 cDNA 测序得到。比较该序列与全基因序列的差异可得知内含子的位置。我们将在后面回顾如何从 mRNA 制备 cDNA。

我们知道当细胞在环境中接受到新的信号时，基因会打开或关闭。DNA 探针与细胞中 RNA 的杂交可用于鉴定特定的基因是否在进行转录。并且，当一个基因的表达量改变时，我们还可以判断出其改变是由于转录水平还是转录后水平的调控（图 7-87）。这些对基因表达的检测最早每次用一个 DNA 探针。如今 DNA 微阵列技术允许人们同时操纵成百上千个基因，我们将在后面讨论。杂交技术在当今生物学研究中是如此重要以至于我们很难想像离开它们如何进行基因结构和表达的研究。

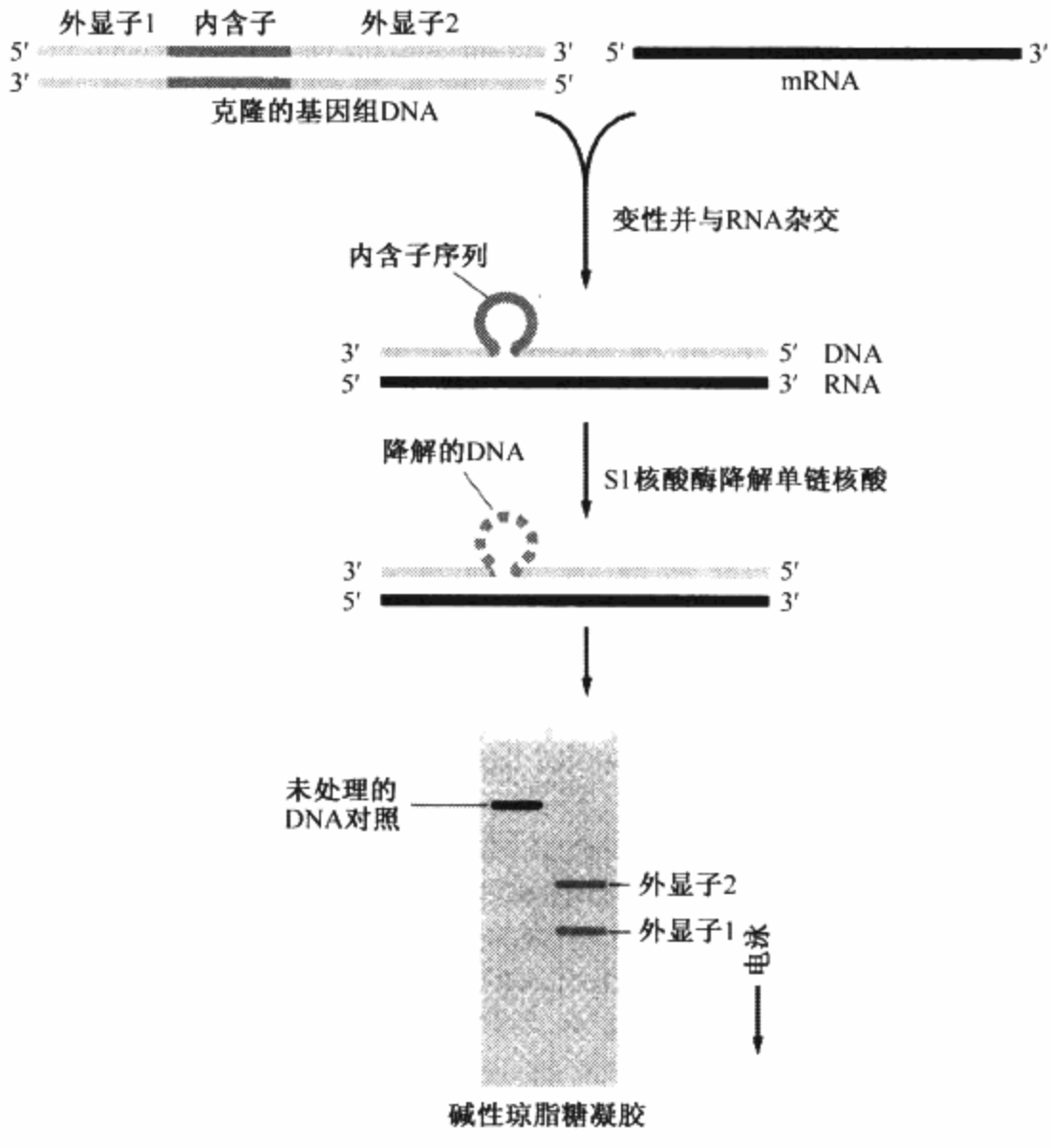


图 8-26　用核酸杂交鉴定 mRNA 分子中出现的克隆的 DNA 片段的区域

所展示的方法中需要一种能够只在与 RNA 链不形成碱基互补配对处切割 DNA 链的核酸酶。真核基因中内含子的位置就是用图中的方法确定的；RNA 分子的起始和终止位置也可以用同样的方法确定。DNA 在变性的琼脂糖凝胶中以单链的形式迁移，结果用于分析

DNA 印迹法、RNA 印迹法和电泳分离的核酸分子的杂交

DNA 探针通常用以在核酸的复杂混合物中检测含有与探针全部或部分序列互补的序列的分子。凝胶电泳可以用于在杂交反应之前依据相对分子质量从粗提的混合物中分离不同的 RNA 和 DNA。如果仅有一种或数种相对分子质量的分子被探针所标记，则可以断定杂交是特异性的。而且，从中得到的相对分子质量信息本身也可能是非常重要的。举一个例子说明。

假设人们想要搞清一种突变小鼠产生缺陷的原因。这种小鼠只能产生异常少量的白蛋白，而通常这种蛋白质由肝脏组织大量产生分泌进入血液。首先，收集突变的和正常的小鼠的肝脏组织样品（后者作为对照），用高强度的去污剂破碎细胞以使核酸酶失活，否则它们将降解核酸。接下来，将 RNA 和 DNA 与其他的细胞组分分离：体系中已经全部变性的蛋白质可用苯酚（一种有效的有机溶剂，可与水部分混合）反复沉淀除去；存在于水相中的核酸可用乙醇沉淀与小分子分离。之后，根据 DNA 与 RNA 在乙醇中的不同溶解性分离 DNA 与 RNA，用高度特异性的酶（RNase 或 DNase）处理可除去残留的不要的组分。通常使用与 mRNA 的多聚腺苷酸尾巴特异性结合的滞留色谱从大量的 RNA 中分离 mRNA。

为了用 DNA 探针寻找编码白蛋白的 mRNA，可以采用一种叫做 **Northern 印迹**的技术。首先，从突变的和对照的肝细胞中纯化出来的全部 mRNA 用凝胶电泳的方法按大小分离成一

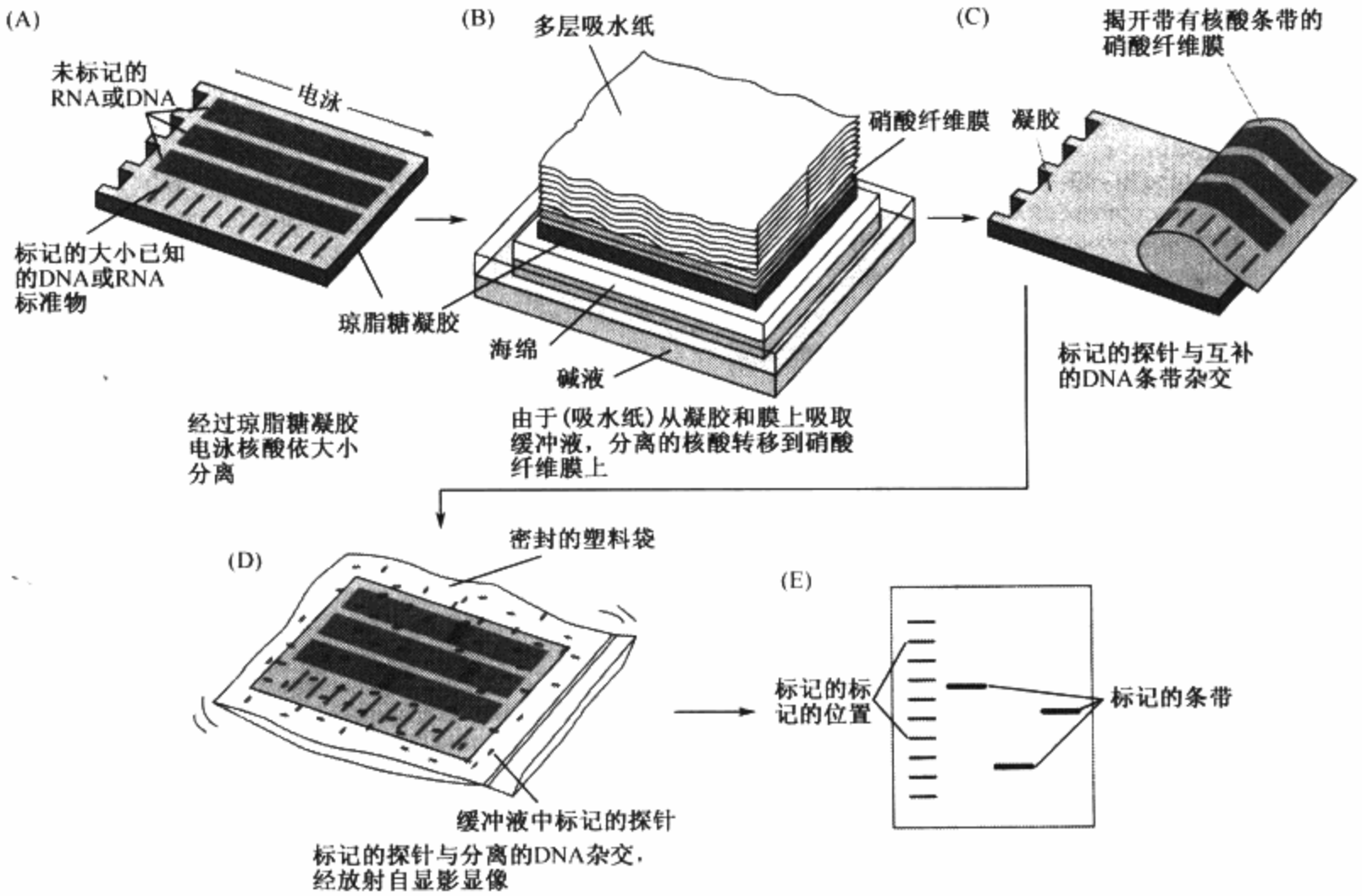


图 8-27　用凝胶转移杂交检测特异性 RNA 或 DNA 分子

在这个例子中，DNA 探针通过其放射性进行检测。用化学或荧光的方法检测 DNA 探针也是很常见的（图 8-24）。（A）单链RNA 分子（Northern Southern 印迹）或者限制酶处理产生的双链 DNA 分子（Southern 印迹）混合物经过电泳分离开。（B）将一张硝酸纤维或者尼龙膜覆于凝胶上，通过印记杂交将 RNA 或者 DNA 转移到膜上。（C）将硝酸纤维膜小心剥离。（D）将含有核酸的膜放入装有缓冲盐溶液和放射性标记的 DNA 探针的密封塑料袋中。在适宜杂交的条件下，膜与放射性 DNA 探针接触足够长的时间。（E）从袋中取出膜，充分洗涤，只有与 RNA 或 DNA 杂交的探针分子结合在膜上不被洗掉。经过放射自显影，与标记的探针杂交的 DNA 显示出条带。在 RNA 印迹法中，膜上的 DNA 双链分子必须在杂交进行前分离。这一过程是在电泳之后把 DNA 置于碱性条件下变性实现的（未显示）

系列的条带。然后，为了使 RNA 分子接触 DNA 探针，通过印迹杂交将 RNA 条带拷贝到硝酸纤维或尼龙膜上。薄膜与含有标记过的 DNA 探针的溶液共浴，探针序列与编码白蛋白 mRNA 的模板链的部分序列互补。膜上与标记的 DNA 探针杂交的 RNA 分子（因为它们与正常白蛋白基因序列部分互补）可通过放射自显影或者化学方法检测（图 8-27）。与探针结合的 RNA 分子的大小可以根据在电泳中置于实验样品旁边的大小已知的 RNA 分子（标准 RNA）的位置进行判断。这样，人们可以知道，突变小鼠的肝细胞是产生与正常肝细胞等量而且大小一样的 RNA 分子还是产生的量有明显的减少。另一种可能是突变的白蛋白 RNA 分子显著变短，因而在胶中移动异常迅速。在这种情况下，可以用一系列更短的 DNA 探针重新检测凝胶，每种探针与基因的一小部分互补，从而揭示出正常 RNA 的哪一部分缺失了。

另一种类似的凝胶转移杂交法叫做 Southern 印迹，是用来分析 DNA 而不是 RNA 的。分离的 DNA 用限制性内切核酸酶切割成可以分离的片段。双链片段依据大小进行凝胶电泳分离，与 DNA 探针互补的片段用与鉴定 RNA（图 8-27）同样的方法进行印记杂交。为了鉴定突变小鼠白蛋白基因的结构，可以采用白蛋白特异性 DNA 探针构建出白蛋白基因区详

细的基因组限制性图谱。从图谱中人们可以看出在无功能的动物中白蛋白基因是否被重新组织了。例如，是否有小片段 DNA 的缺失或插入，但是多数的单碱基突变用这种方法不能判定。

用杂交技术可以在细胞或者染色体上对特定的核酸序列进行定位

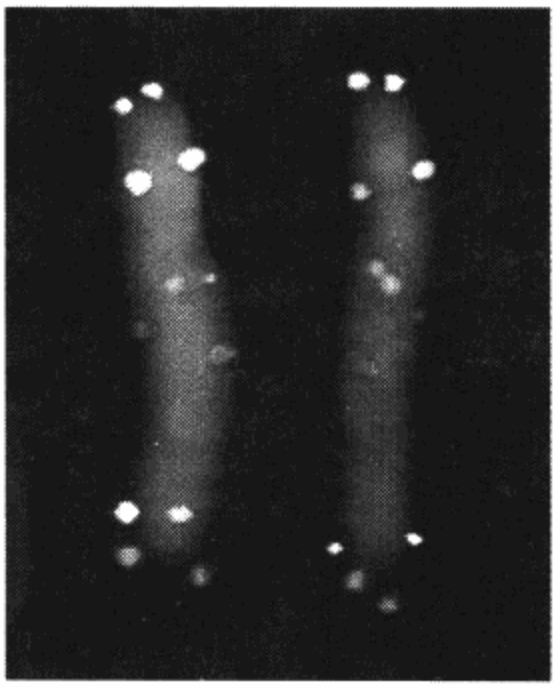


图 8-28　通过原位杂交在染色体上对特定基因进行定位

在这里，6 个不同的 DNA 探针用于在有丝分裂中期的人类 5 号染色体上对它们各自对应的核酸序列进行定位。探针已进行化学标记，可以用带荧光的抗体检测。两条 5 号染色体并排展示在上面。每种探针在每条染色体上显示出两个亮点，因为分裂中期的染色体经过复制具有两个相同的 DNA 双螺旋（经 David C. Ward 许可）

与其他大分子一样，核酸在细胞和组织中也有精确的定位，并且当这些分子经过纯化提取后，其带有的大部分潜在信息都丢失了。所以，人们发明了与用标记的抗体原位定位特定核酸序列类似的方法用核酸探针进行**原位杂交**。这项技术现在既可以用于在染色体上定位 DNA，也可以用于在细胞中定位 RNA。经过短时间在极高的 pH 下，染色体的 DNA 碱基对被打开，可以与标记的核酸探针进行杂交。与探针杂交的染色体区域可以看到。开始这项技术采用的是高放射性的 DNA 探针，通过放射自显影显像。但是后来人们采用前面化学标记的方法（图 8-28）代替了已经介绍过的放射性的手段。

原位杂交方法的发展也帮助人们揭示了组织细胞中 RNA 的分配情况。在这种情况下，不用高 pH 处理组织，所以染色体 DNA 是成双链的，不与探针结合。把组织经过温和的固定之后，RNA 处于暴露的状态，经过温浴可以与互补的 DNA 或 RNA 探针杂交。这样，可以在组织中观察差异性的基因表达，还可以确定特定 RNA 在细胞中的位置（图 8-29）。例如，这种方法为人们提供了研究果蝇中不同位置的细胞在发育中产生差异的机制的途径（讨论见第 21 章）。

可以从 DNA 文库中克隆基因

每个人们所感兴趣的 DNA 片段都可以被克隆。在细胞生物学中，**DNA 克隆**有两种含义。一种含义是指它的表面意思，即得到一个 DNA 分子许多相同的拷贝的行为——扩增一个特定的 DNA 序列。而这个术语的另一种含义指从细胞中分离一段特定的 DNA（通常是一个特定基因），因为得到特定 DNA 的大量拷贝非常有助于分离它。

在这个最普遍使用的含义上，进行 DNA 克隆有很多方法。最简单的方法是将一段特定的 DNA 片段插入到一个自主复制基因元件纯化的基因组中——通常是一个病毒或者一个质粒。例如，一个含有人类基因的 DNA 片段可以在试管中与噬菌体的染色体接在一起，重组得到的 DNA 分子可以引入细菌细胞中。从一个感染细胞的重组 DNA 分子，经过一般的病毒复制装置可以在一天之内得到 10^{12} 个以上相同的病毒 DNA 分子，同时也将人类的 DNA 片段放大了同样的倍数。一个这样的病毒或质粒叫做一个克隆载体，插入其中得以扩增的 DNA 就叫做被克隆了。

要分离一个基因，人们通常从构建 DNA 文库开始，即从细胞、组织或者器官中克隆的 DNA 片段的总集。人们希望这个文库至少含有一个带有目的基因的片段。文库通常用病毒或者质粒构建，然后以细菌细胞为宿主。尽管细节可能有所区别，但是两种克隆方法的原理是相同的。今天大多数克隆倾向于选择质粒载体。

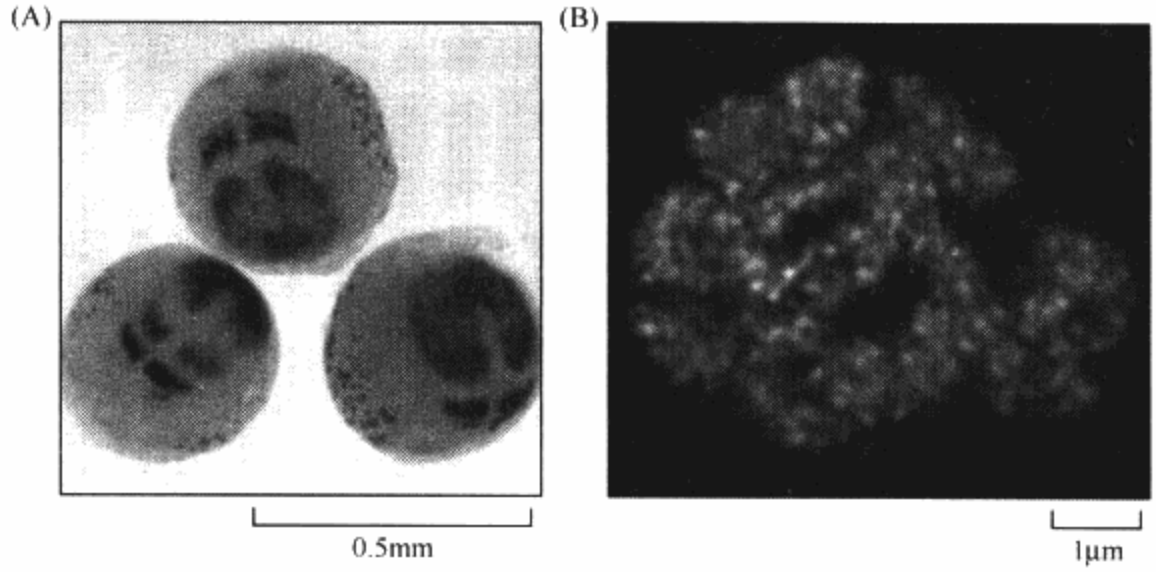


图 8-29　用原位杂交定位 RNA

（A）斑马鱼早期胚胎 *deltaC* 的表达模式。这个基因编码一个 Notch 信号途径中的配体（讨论见第 15 章），显示的模式反映了其在体节（脊椎动物未来的躯干和尾巴）发育中的作用。（B）高分辨率的 RNA 原位杂交展示出豌豆细胞中核糖体 RNA 合成的位置。直径 0.5～1μm，香肠状的结构对应于含有编码 rRNA 的染色体 DNA 环。每一个白色小点代表一个正在转录的 rRNA 基因［（A）经 Yun-Jin Jiang 许可；（B）经 Peter Shaw 许可］

最常用于基因克隆的**质粒载体**是从在细菌细胞中正常存在的大质粒改造而来的小型环状双链 DNA 分子。它们通常只占宿主细菌细胞总 DNA 的一小部分，但却因为相对分子质量小很容易与染色体 DNA 分离，因为它们不像巨大的染色体 DNA 那样在离心的时候形成颗粒沉淀。作为克隆的载体，纯化的质粒 DNA 环用限制性内切核酸酶切成线性分子。用于构建文库的细胞 DNA 也用同样的限制酶酶切，得到的限制性片段（包含那些含有待克隆基因的）加入到酶切的质粒中，通过它们的黏性末端退火形成重组 DNA 环。其后，用 DNA 连接酶将这些含有外源 DNA 插入片段的重组分子进行共价连接（图 8-30）。

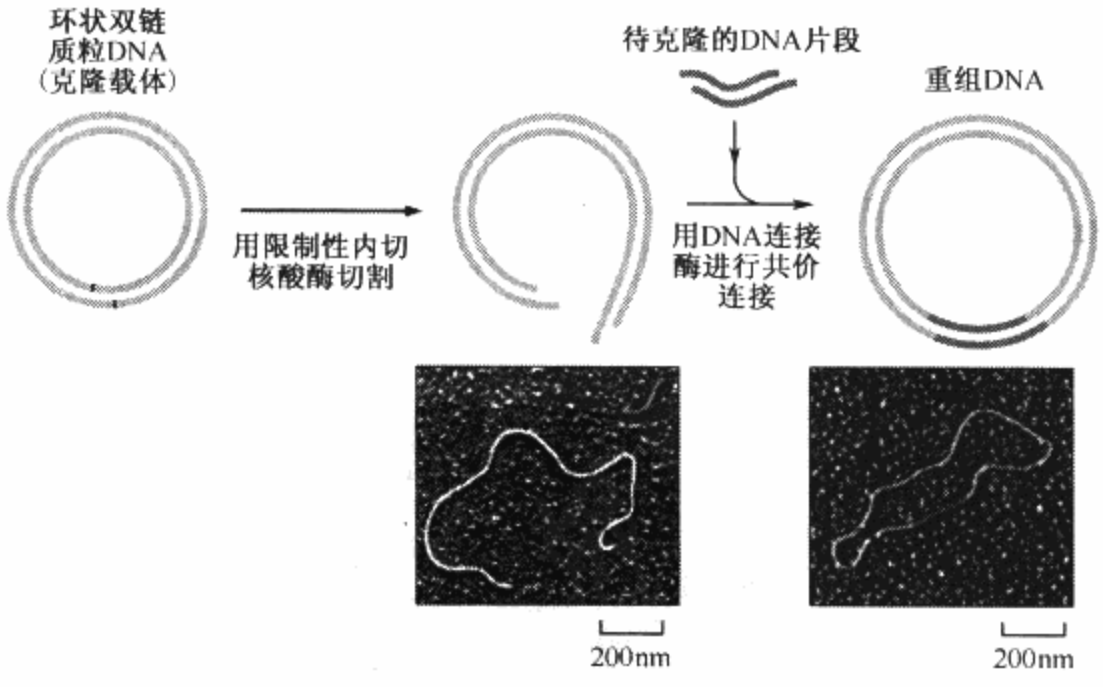


图 8-30　用 DNA 连接酶将一个 DNA 片段插入细菌质粒载体

用限制性内切核酸酶切开质粒（此时产生黏性末端），与待克隆的 DNA 片段（也用相同的限制性内切核酸酶处理）、DNA 连接酶和 ATP 混合。黏性末端形成碱基配对，DNA 连接酶填平 DNA 主干上的缺口，产生完整的重组 DNA 分子（显微摄影经 Huntington Potter 与 David Dressler 许可）

构建文库的下一个步骤是将重组 DNA 环导入对 DNA 短时可通透的细菌细胞中，称作用质粒转染细胞。随着这些细胞的生长和每 30min 繁殖一代，含有外源 DNA 的 DNA 环也进行

复制，产生大量拷贝（图 8-31）。很多细菌质粒携有抗生素抗性基因，可以用来筛选转染成功的细胞。在有抗生素存在的条件下，只有含有质粒的细胞才能生存。通常每个初始转染的原始细菌细胞中含有一个彼此不同的外源插入 DNA。该 DNA 被其所有后代继承，并共同形成了培养皿中的一个菌落。

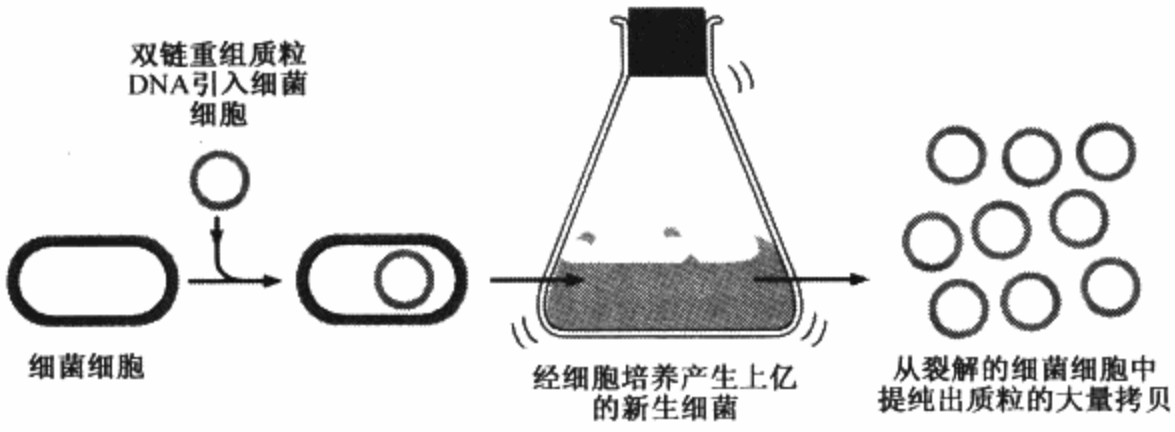


图 8-31　通过细菌中的 DNA 克隆纯化和扩增特定 DNA 序列

要得到一个特定 DNA 序列的大量拷贝，先要把片段插入到质粒载体中，如图 8-30 所示。之后，重组质粒 DNA 被引入细菌中，随着细菌的繁殖而扩增百万倍

很长时间里，质粒被用于克隆 1000～3000bp 长度的 DNA。更长的 DNA 片段更难以操纵和克隆。因此研究者们转向使用酵母人造染色体（YAC），它可以操纵很长的 DNA 片段

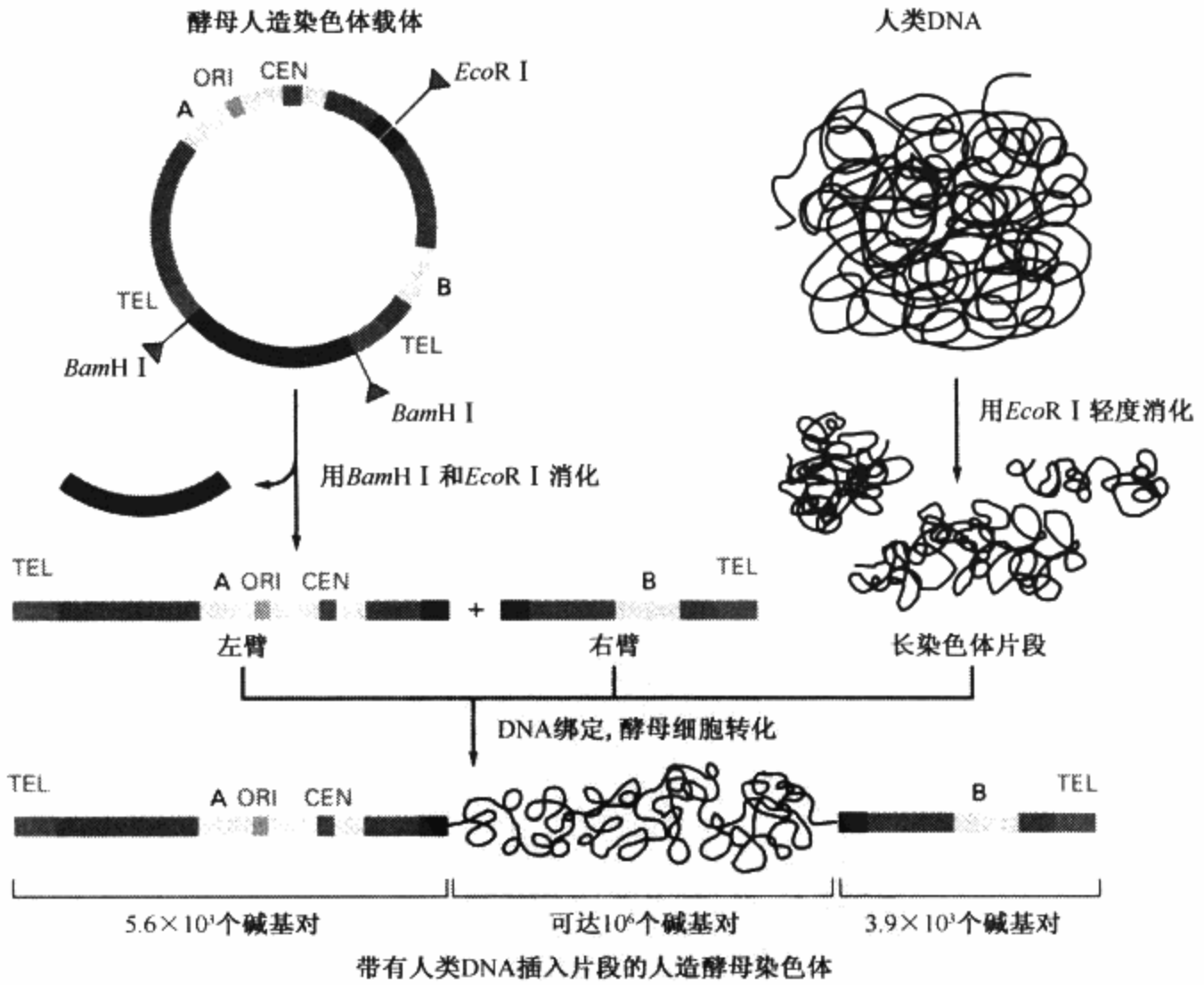


图 8-32　制造酵母人造染色体（YAC）

用 YAC 载体可以克隆很大的 DNA 分子。TEL、CEN 和 ORI 分别代表酵母的端粒、着丝粒和复制序列起点。*Bam*H Ⅰ和 *Eco*R Ⅰ代表相应的限制性内切核酸酶在 DNA 双螺旋上的切割位点。标有 A 和 B 的序列编码作为选择标记的酶，可以简单地分离出携有人工染色体的酵母细胞（摘自 D. T. Burke, G. F. Carle, and M. V. Olson, *Science* 236：806～812，1987）

(图 8-32)。现在，用从天然的大肠杆菌 F 质粒改造而来的新质粒可以克隆长度在 300 000bp 到 100 万 bp 之间的 DNA 片段。与小细菌质粒不同，F 质粒及其衍生物**细菌人造染色体**（BAC）在大肠杆菌中只有一两个拷贝。因为拷贝数低，所以它们倾向于稳定保存大的克隆片段，而很少引起质粒拷贝上的片段发生重组。由于它们的稳定性和对大 DNA 插入片段的容纳性，以及易于操纵的性质，现在 BAC 是构建复杂生物，包括人类和小鼠基因组中表达的 DNA 文库的首选载体。

两种不同的 DNA 文库用于不同的目的

用一种特定的限制性内切核酸酶切割全基因组，并用上面描述的方法将每一个片段克隆的克隆法有时候叫做“鸟枪法”。用这项技术可以得到数目巨大的 DNA 片段（对于哺乳动物基因组可达百万量级）从而产生百万个不同的转染菌落。（当使用 BAC 代替一般的质粒时，可以插入较大的片段，所以覆盖基因组所需的转染细菌细胞数目较少。）每一个这样的菌落都是由一个祖细胞繁衍而来的，因此携带有片段化了的基因组某一部分的许多拷贝（图 8-33）。这样的一个质粒叫做含有一个**基因组克隆**，全部质粒的集合叫做一个**基因组 DNA 文库**。但是因为基因组 DNA 是随机片段化的，所以只有某些片段含有基因。很多从高等真核生物中得来的基因组 DNA 克隆只含有非编码 DNA，它们组成了基因组 DNA 的绝大部分，正如我们在第 4 章中讨论过的那样。

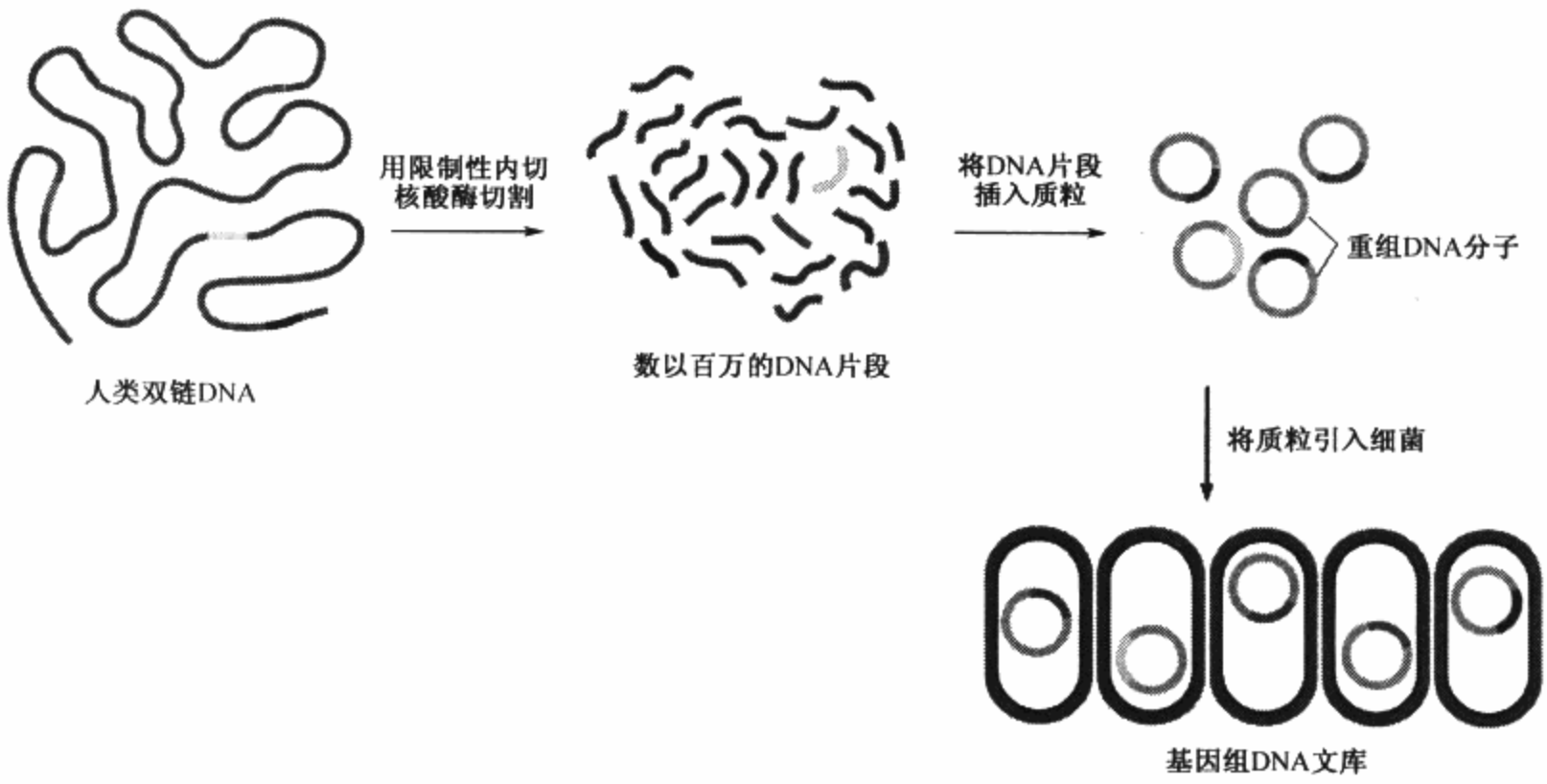


图 8-33　构建人类基因组 DNA 文库

一个基因组文库通常保存在一系列的细菌中，每个菌带有一个人类 DNA 的不同片段。图中简单表示出了几个示意的片段（有颜色的）。实际上，所有的灰色 DNA 片段都要被克隆

另一种策略是仅选择转录成 mRNA 的 DNA 序列进行克隆，它们对应于蛋白质编码基因。按照这种策略，首先要从细胞中提取 mRNA（或纯化 mRNA 组分），然后产生每一个 mRNA 分子互补的 DNA 拷贝（cDNA）。这个反应是通过反转录病毒的反转录酶实现的，它可以以 RNA 链为模板合成 DNA 链。DNA 聚合酶可以将单链 DNA 分子转化合成双链 DNA 分子，其后插入质粒或病毒载体进行克隆（图 8-34）。每一个这样得到的克隆叫做一个 **cDNA 克隆**，从制备的 mRNA 得到的全部克隆叫做一个 **cDNA 文库**。

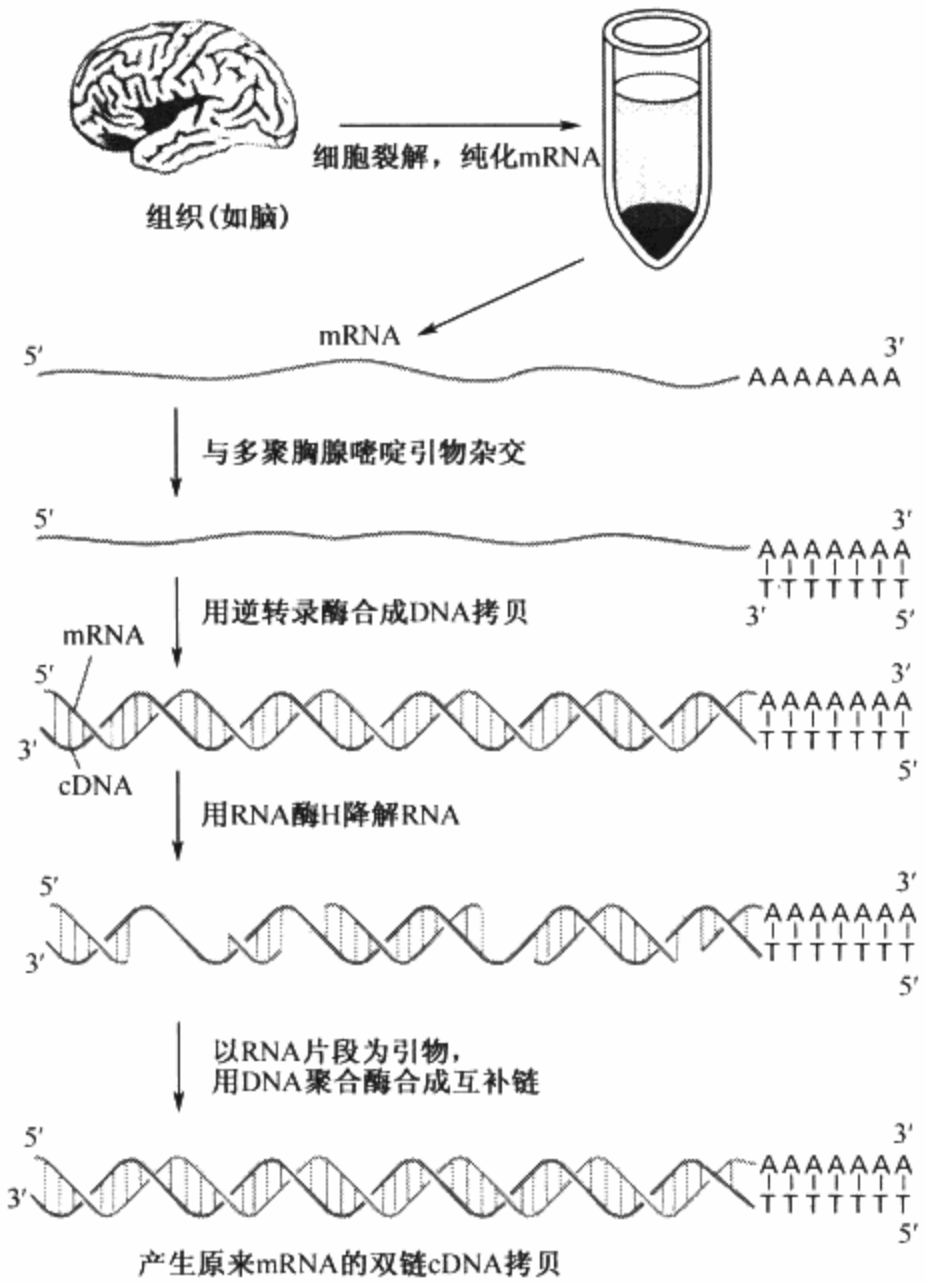


图 8-34　cDNA 的合成

从特定组织中提取总 mRNA，通过反转录得到 mRNA 分子的 DNA 拷贝（cDNA）。为了简化，只展示了一个 mRNA 拷贝成 DNA 的过程。与 mRNA 的 3′端多聚腺苷酸链（讨论见第 6 章）互补的短的寡核苷酸链先与 RNA 杂交作为反转录的引物，之后形成 RNA 链的互补 DNA 链，从而形成 DNA/RNA 杂合双螺旋。用 RNA 酶 H（图 5-13）处理杂合链，在 RNA 链上产生缺口。剩下的单链 cDNA 经 DNA 聚合酶拷贝成双链 cDNA。如图所示，原来 RNA 的一个片段作为这一反应的引物。因为合成第二条链的 DNA 聚合酶在合成的时候会跨过结合的 RNA 分子，所以与第一条 DNA 链的 3′端形成碱基互补 RNA 片段通常作为第二条链合成终产物的引物。这条 RNA 链在其后的克隆步骤中被降解。所以，原来 mRNA 的 5′端序列总是在 cDNA 文库中缺失

基因组 DNA 克隆与 cDNA 克隆之间有很大的区别，如图 8-35 所示。基因组克隆展示了从一个生物体中得到的所有随机样品的 DNA 序列，除了极少的例外，对于不同的细胞类型是相同的。相反，cDNA 克隆只含有被转录成 mRNA 的基因组区域。因为不同组织的细胞产生不同的 mRNA 分子，所以每一种不同的细胞类型产生不同的 cDNA 文库。

cDNA 克隆含有不间断的编码序列

用 cDNA 文库进行基因克隆有很多优点。首先，许多蛋白质在分化的细胞中产量很高。在这种情况下，从细胞制备的编码该蛋白质的 mRNA 就有足够多的量产生大量的 cDNA，从而大大降低在文库中寻找所需基因的工作量（图 8-35）。例如，红血球（红细胞）可以产生大量的血红蛋白，因此血红蛋白基因是最先被克隆的基因之一。

cDNA 克隆最大的优点是它们含有连续的基因编码序列。如我们所知，真核基因通常是

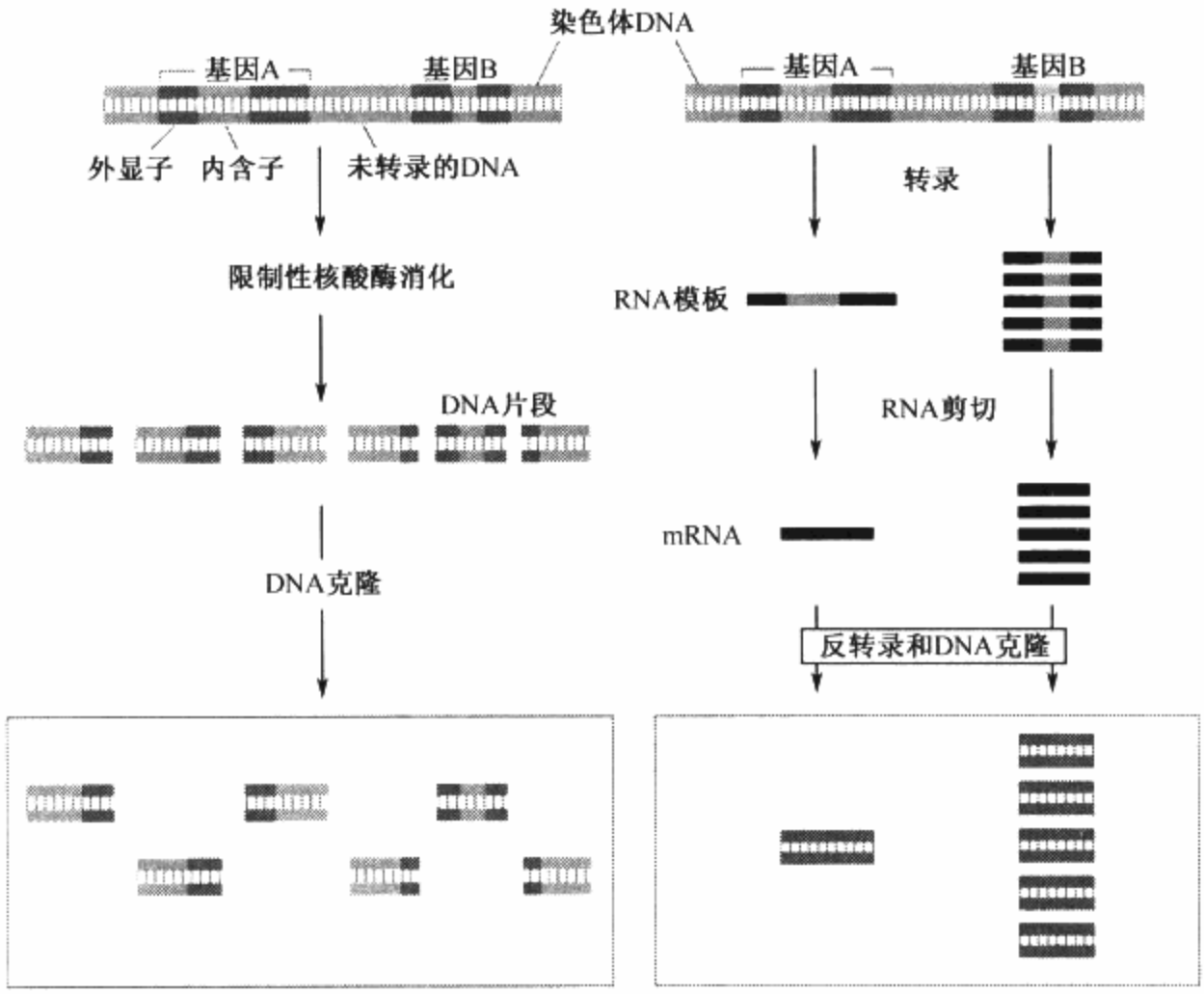


图 8-35　一个 DNA 区域的 cDNA 克隆与基因组 DNA 克隆之间的区别

在这个例子中，基因 A 的转录频率低，基因 B 的转录频率高，两个基因都含有内含子（绿色表示）。在基因组 DNA 文库中，内含子和非转录 DNA（用粉色表示）都被克隆，而大部分克隆中都只含有基因编码序列（红色表示）的一部分。在 cDNA 克隆中，内含子序列（黄色表示）在mRNA（蓝色表示）形成过程中被剪切掉，因此每个克隆中出现的是一个连续的编码序列。因为在构建文库所用的细胞中基因 B 的转录频率比基因 A 高，所以它在 cDNA 文库中出现的次数也比 A 高。与此相反，A 与 B 在基因组 DNA 文库中出现的次数大体相当

由被很长的非编码序列（内含子）隔断的较短编码序列（外显子）组成的，需要从初生 RNA 上移除非编码序列并将编码序列接在一起时才能产生 mRNA。细菌和酵母细胞都不需要高等真核生物产生 RNA 的修饰过程。所以，当克隆的目的是从 DNA 序列推导蛋白质的氨基酸序列或者在细菌或酵母中表达克隆的基因以得到大量蛋白质时，人们更倾向于选择从 cDNA 开始实验。

基因组 cDNA 文库是研究者共同享有的取之不尽的资源。今天，许多这样的文库也可以通过商业渠道获得。

分离的 DNA 片段能够进行快速测序

20 世纪 70 年代晚期，人们发明了能够简单而迅速地对任何纯化的 DNA 片段进行核苷酸测序的方法。它们使得测定数万个基因的完整 DNA 序列成为可能，许多生物的基因组序列也得以测定出来（表 1-1）。关于 DNA 序列的信息是如此浩繁（数百亿个核苷酸）以至于需要使用功能强大的计算机去储存和分析它们。

DNA 序列的大量测定成为可能是从 20 世纪 70 年代中期**双脱氧法**测定 DNA 序列开始的。该方法基于在链终止双脱氧核糖核苷三磷酸存在状况下的体外 DNA 合成（图 8-36）。

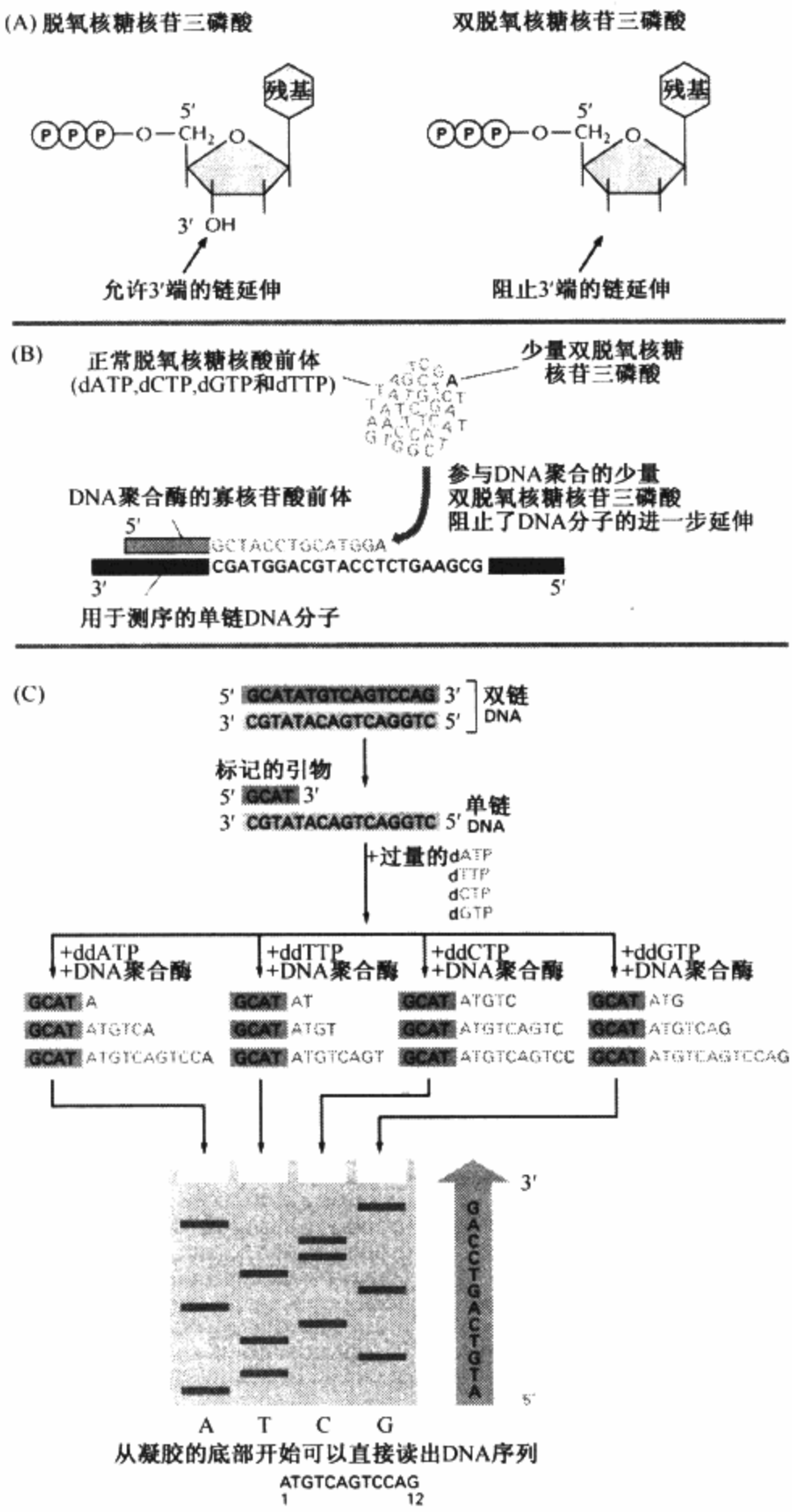


图 8-36　酶法（即双脱氧法）测定 DNA 序列

（A）这种方法依赖于从正常的脱氧核糖核苷三磷酸衍生而来的缺少 3′-羟基基团的双脱氧核糖核苷三磷酸的使用。（B）在含有单链的 DNA 待测序列（灰色）、DNA 聚合酶、引发聚合酶合成开始的短 DNA 引物（橙色）以及 4 种脱氧核糖核苷三磷酸（dATP、dTTP、dCTP 和 dGTP，用绿色的 A、C、G、T 表示）的体外混合物中，合成纯的 DNA。如果一种核苷酸的双脱氧核糖核苷三磷酸类似物（红色表示）也存在于混合体系中的话，它也可以参与链的合成。因为这样的链缺少 3′-羟基，所以不能进一步加入新的核苷酸而在该点产生链终止。在示例中，核苷酸混合物中含有少量的双脱氧 ATP（ddATP，此处用红色的 A 表示）。它们与过量的正常脱氧 ATP（dATP，用绿色的 A 表示）竞争，随机参与 DNA 链的延伸。反应混合物最终产生一系列不同长度的与模板 DNA 互补的 DNA，分别在延长链的不同的 A 处发生终止。（C）为了鉴定完整的 DNA 片段的序列，首先将双链 DNA 分成 2 个单链，其中一个作为测序的模板。4 种不同的链终止双脱氧核糖核苷三磷酸（ddATP、ddCTP、ddGTP、ddTTP，仍然用红色表示）用于模板（用灰色表示）相同的 4 个分开的 DNA 合成反应。反应产物在 4 个并排的聚丙烯酰胺凝胶泳道中电泳分离（用 A、T、C、G 表示）。新合成的片段用连在新生链中的引物或一种脱氧核糖核苷三磷酸作为标记（放射性或荧光的）进行检测。在每一个泳道中，条带代表在给定的核苷酸处（如最左边则是 A）终止但终止位置不同的核酸链。按照从下到上的顺序在各条泳道中读出条带，可以确定新生 DNA 的序列。序列在凝胶右侧的肩头中给出。这个序列与原来的双链 DNA 分子中 5′→3′链（绿色）的序列相同

尽管基本的方法今天仍在使用，但许多地方已经进行了改进。DNA 测序如今已经完全自动化：机器手混合试剂、加样、电泳并读出核苷酸碱基序列。通过对链终止核苷酸进行不同颜色的荧光染料标记可以提高效率，而且在这种情况下，4 个反应全都可以在同一个试管中进行，产物也可以在同一个凝胶泳道中进行分离。位于凝胶底部的检测器在激光通过电泳条带的时候可以记录下每个条带的荧光颜色（图 8-37）。计算机可以读出并记录下核苷酸的序列。

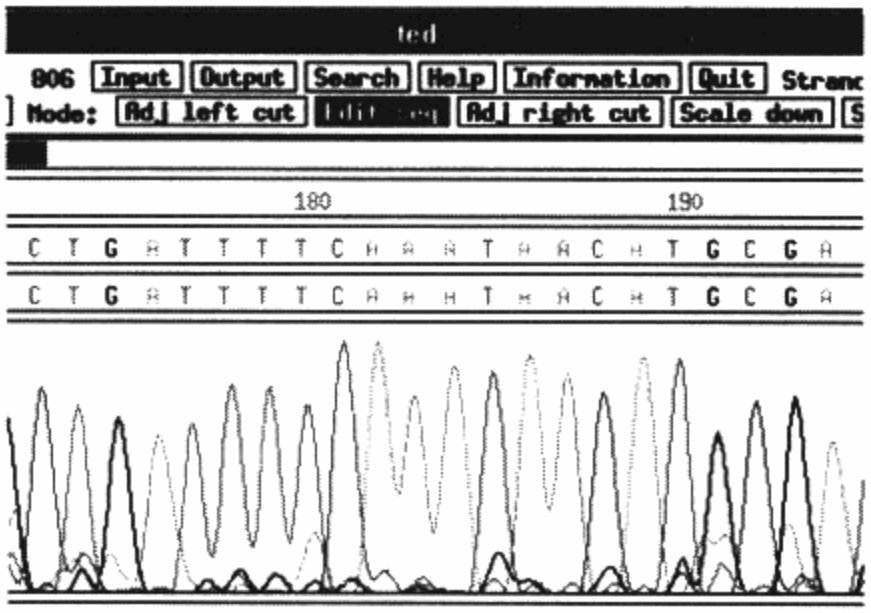


图 8-37　自动 DNA 测序

这里展示的是 DNA 自动测序过程中计算机所显示的一小部分数据。每一个有颜色的峰代表 DNA 序列中的一种核苷酸，从序列起点 173 到 194 位的核苷酸序列可以清楚地读出来。这个例子来自拟南芥全基因组的国际项目（经 George Murphy 许可）

核酸序列用于预测蛋白质的氨基酸序列

今天 DNA 测序是如此的快速可靠以至于它已经成为间接判定大多数蛋白质的氨基酸序列的首选方法。有了编码一个蛋白质的核酸序列之后，步骤就十分明了。尽管在理论上有 6

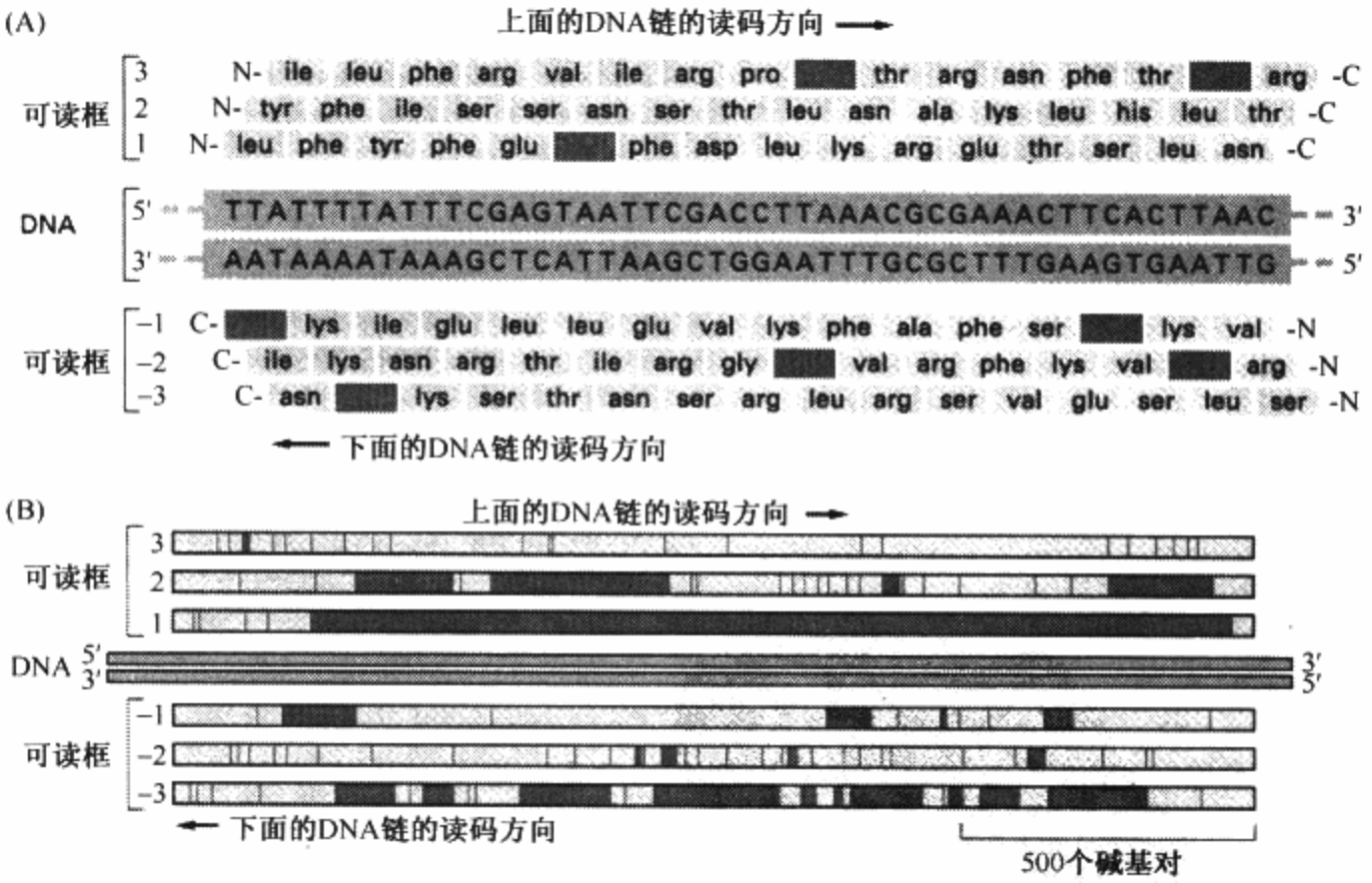


图 8-38　寻找编码蛋白质的 DNA 序列区域

（A）任一个 DNA 序列区域基本上都可以编码 6 个氨基酸序列，因为每条链上都有 3 个不同的可读框可用于核酸序列的翻译。注意核酸序列总是以从 5′到 3′的顺序编码从氨基（N）端到羧基（C）端的多肽链。用一个特定的可读框读一个随机的核酸序列时，平均每读 21 个氨基酸（即 63 个核苷酸）就会遇到一个终止信号。在这个 48bp 的例子中，每个终止子都用绿色表示，只有可读框 2 没有终止信号。（B）在 1 700bp 的 DNA 序列中搜索可能的蛋白质编码序列。过程与（A）中一样，每个蛋白质合成的终止信号用绿线表示。除此之外，所有蛋白质合成可能的起始和终止位点之间的区域都用红色的条框表示。只有可读框 1 实际上编码一条 475 个氨基酸长度的蛋白质

个翻译蛋白质的可读框，但是通常唯一正确的可读框是可以从缺少频繁出现的终止子来判断的（图 8-38）。我们在第 6 章中已经讨论过，随机的核酸序列在读码时每 20 个氨基酸左右会出现一个蛋白质合成的终止信号。编码比这长得多的氨基酸序列的核酸序列可以看作是可能的外显子。可以把它们翻译成氨基酸序列（用计算机），并在数据库中搜索其他生物中与之具有相似性的蛋白质。如果需要的话，可以对纯化的蛋白质进行部分测序以验证从 DNA 预测的序列的正确性。

问题在于，如何在整个基因组中确定编码蛋白质的基因。在缺少内含子的细菌和古细菌染色体中或者 cDNA 克隆中确定基因是最容易的。通过对 DNA 某些明显的性质进行鉴定（在第 6 章中已讨论），可以推测基因在这些核酸序列中的定位。简而言之，编码蛋白质的基因是通过通常以 ATG 为起始密码子，以 TAA、TAG 或 TGA 为终止密码子的可读框（ORF）鉴定的。为了减少错误，通常要求计算机只考虑长度大于 100 个密码子的序列。

对于更复杂的基因组，如真核生物基因组，这个过程被嵌在基因的编码成分之中的内含子复杂化了。在包括人在内的许多细胞生物中，外显子的平均长度只有 150bp。所以在真核生物中，人们必须寻找基因存在的其他特征，如内含子/外显子边界或者明显的上游调节区域。

另一种判断染色体中编码区的方法是通过鉴定可检测到的 mRNA 的核酸序列（以 cDNA 的形式）的性质。mRNA（以及由它们产生的 cDNA）缺少内含子、调节 DNA 序列和基因之间非必需的“间隔”DNA。这对于大量测定 cDNA 序列以得到一个某一生物编码序列的巨大数据集合（叫做一个数据库）是有帮助的。之后这些序列可以用于在与基因对应的长的染色体 DNA 序列中区别内含子与外显子。

最后，有亲缘关系的生物中的保守核酸序列通常编码蛋白质。在不同的生物中比较这些序列可以提供有关一个特定蛋白质或基因的功能的信息，我们将在下一章讨论。

许多生物的基因组都全部测序完成

受益于 DNA 测序的自动化，许多生物的基因组已经完全测序完成了，包括植物叶绿体、动物的线粒体、为数众多的细菌和古细菌，以及许多在实验室中广泛研究的模式生物，包括几种酵母、一种线虫、果蝇、模式植物拟南芥、小鼠，还有人类。研究者已经得到了大量的人类病原体的完整 DNA 序列，包括霍乱、肺结核、梅毒、淋病、莱姆氏病和胃溃疡，还包括数以百计的病毒，如天花病毒以及爱普斯坦巴氏小体病毒（能引起传染性单核细胞增多症）。对这些病原体基因组的研究可能会提供关于它们如何致病的线索，从而指出更有效的治疗新途径。

流感病菌（一种引起儿童耳部感染或脑膜炎的细菌）是第一个全基因组测序完成的生物（共 180 万个核苷酸），其测序采用了今天最广泛采用的鸟枪测序法。使用鸟枪法，DNA 的长序列被随机地打成较小的片段，每个片段分别测序，再利用片段间的重叠部分用计算机将其组合成完整的染色体或基因组。鸟枪法是一种小基因组测序技术。尽管高重复性的基因组用鸟枪法测序会比较麻烦，在果蝇、小鼠和人类基因组测序中鸟枪法仍然十分有用。

随着新的序列在科学文献中以不断加快的速度出现，对不同生物的全基因组序列进行比较使我们得以追踪这些基因和生物的进化关系，并发现基因和预测它们的功能。确定基因的功能通常涉及将它们的序列与在实验室中已充分研究的模式生物，如大肠杆菌、裂殖酵母、秀丽新小杆线虫和果蝇（在第 1 章中已讨论）的相应序列进行比对。

尽管基因组已全部测序完成的生物有很多相同的细胞途径，并且在许多蛋白质的氨基酸在序列和结构上具有同源性，但是最新发现的大量蛋白质的功能仍然不得而知。基因组编码的 15%～40%的蛋白质与任何功能已知的蛋白质并不相似。这个现象说明了基因组领域研究

存在的一个局限性：尽管基因组的比较分析提供了有关基因与生物体关系的很多信息，但是通常不能直接解释这些基因如何行使功能，或者说明它们在生物体的生理状况下存在哪些规律。例如，通过若干嗜热细菌全基因组的比较，仍无法得知它们为什么能够在超过 70℃的高温下生存。对神奇的抗辐射细菌基因组的研究并不能解释为什么这种生物能够在足以粉碎玻璃的射线下存活。为了了解在活的生物体中基因如何行使功能，需要进一步进行生物化学和遗传学研究，我们将在本章的最后几节中介绍。

选择的基因片段可以经聚合酶链反应在试管中克隆

既然如此多的基因组序列都已经得到，那么我们就可以不建立 DNA 文库而直接进行基因克隆。**聚合酶链反应（PCR）**技术使快速克隆成为可能。采用 PCR 可以将一个基因组的选定区域扩增数 10 亿倍，从而有效地将这段 DNA 从基因组中纯化出来。

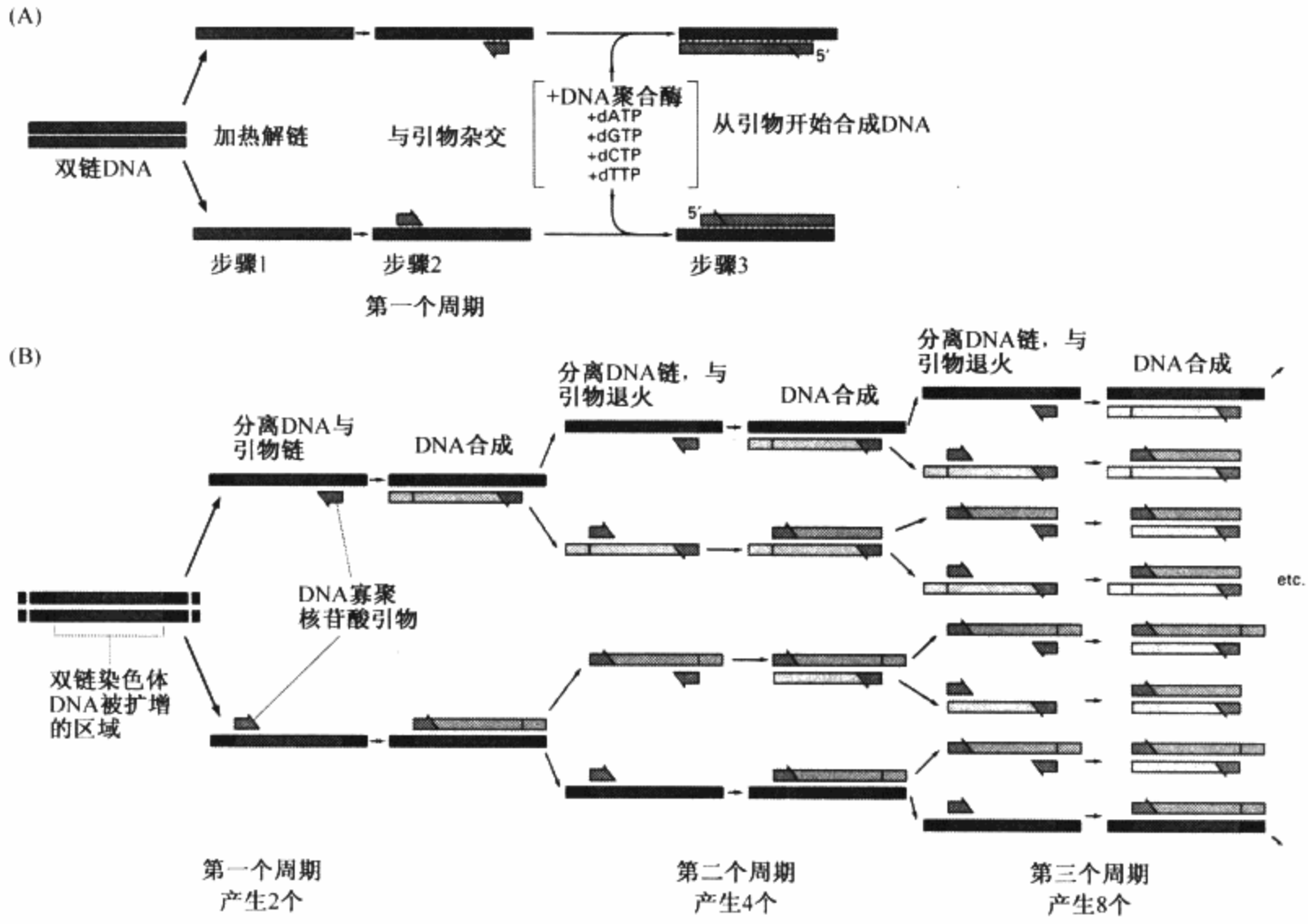


图 8-39　用 PCR 技术扩增 DNA

根据已知的 DNA 序列信息合成两条寡核苷酸链，分别与 DNA 双螺旋的两条链上要克隆区域的一端互补。这些寡核苷酸作为引物与扩增的 DNA 片段识别，在 DNA 聚合酶的作用下，用于体外的 DNA 合成。（A）PCR 以双链 DNA 起始，在每个反应周期中先进行短暂的热处理，使双链 DNA 分开（步骤 1）。解链后，在两个 DNA 寡聚核苷酸引物过量存在的情况下冷却 DNA，引物与两条 DNA 链中的互补序列杂交（步骤 2）。混合物与 DNA 聚合酶及 4 种脱氧核糖核苷三磷酸共浴从引物处开始合成 DNA（步骤 3）。整个周期经加热解链后再次重复。（B）随着过程的不断重复，新合成的片段其后成为模板，在数个周期之内原始模板两个引物之间的序列成为主要的产物。在参与最初反应的 DNA 中，只有位于两个引物之间的序列才能被扩增，因为引物没有其他的结合位点。在（B）例子中，经过 3 个周期产生 16 条 DNA 链，其中 8 条（黄色标志）是与最左边原始标框中序列精确互补的序列；其他的链中含有在起始的几个周期中复制产生的原始序列下游额外的 DNA。再经过 3 个周期之后，256 条 DNA 链中有 240 条链与原始标框中的序列精确互补，再经过几个周期后，几乎所有的 DNA 链都具有相同的长度

为了界定需要的基因序列，我们用化学方法合成两套寡核苷酸，作为在全基因组经加热后得到的单链上合成 DNA 的引物。在体外纯化得到的 DNA 聚合酶的催化下合成新的 DNA，引物留在最终形成的 DNA 片段的 5′端［图 8-39(A)］。

在最初的 DNA 合成周期中，并没有什么特别的地方，但在经过若干周期之后，PCR 的优势就显示出来。因为在每一个周期中需要用短暂的热处理将模板 DNA 双螺旋分成两条单链，所以需要一种从嗜热菌中分离出来的，能够耐受超常温度不致变性的特殊的 DNA 聚合酶。在每一个 DNA 合成周期中新生的片段都可以在下一周期中作为模板，所以经过几个周期之后，产物中的主要成分就成了与两套原始引物之间的长度一致的单链 DNA 片段［图8-39(B)］。

在实践中，需要 20～30 个反应周期来获得足够的 DNA，因为每一周期中的产物都可以作为下一周期的模板，所以得名聚合酶链反应。一个周期只需要 5min，并且整个过程很容易

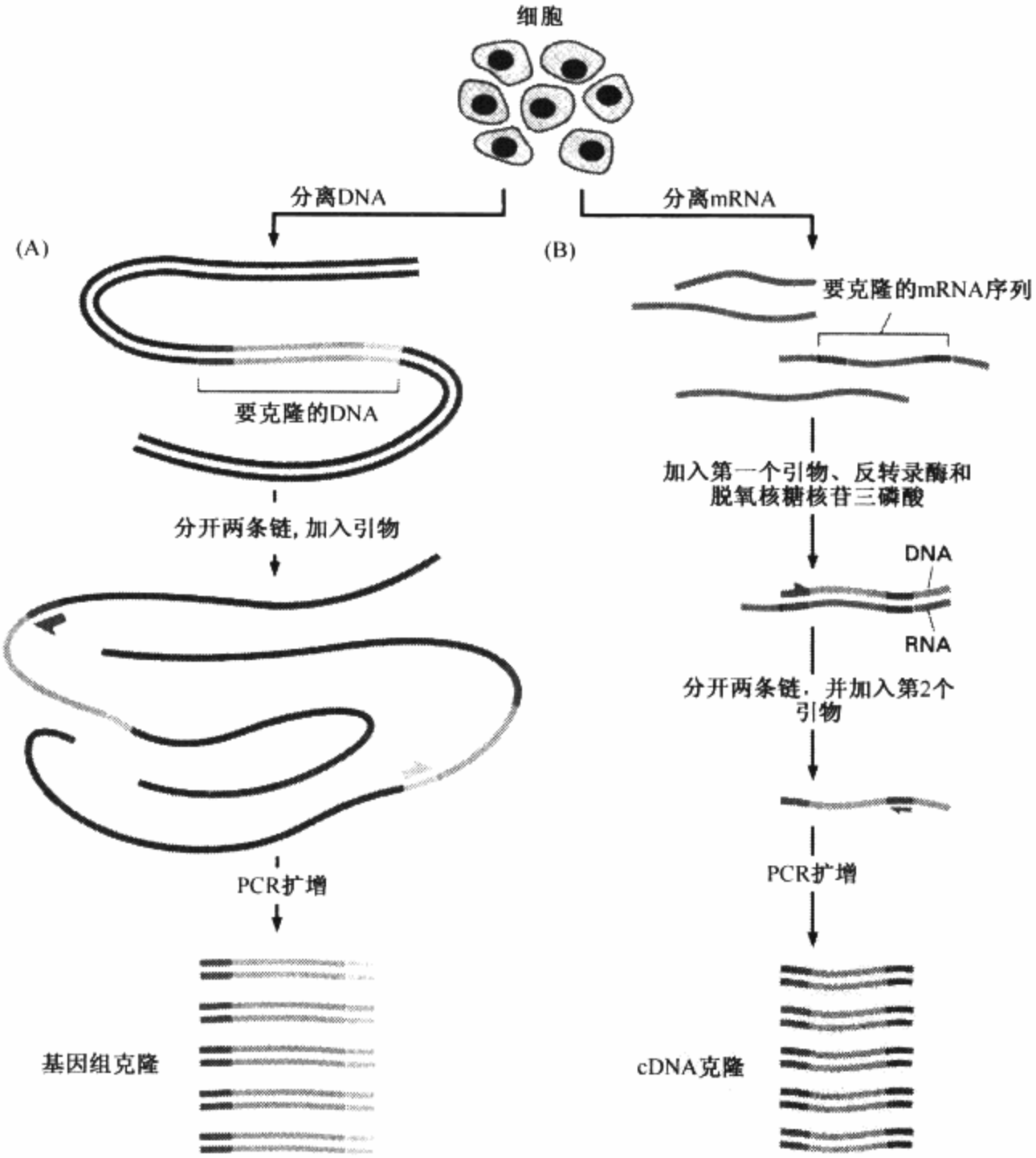


图 8-40　用 PCR 获得基因组或 cDNA 克隆

（A）为了用 PCR 获得基因组克隆，首先从细胞中纯化染色体 DNA。加入位于欲克隆的 DNA 两端的引物，进行多个周期的反应（图 8-39）。因为只有引物之间（包括之）的 DNA 被扩增，所以 PCR 提供了一种从染色体 DNA 选择性获得纯的短链 DNA 的方法。（B）用 PCR 的方法获得 cDNA，首先要从细胞中纯化 mRNA。之后将第一个引物加入 mRNA 中，用反转录酶产生互补的 DNA 链。然后加入第二个引物，经过多个 PCR 反应周期得到单链的 DNA 分子，如图 8-39 所示。对于两种克隆，要克隆的区域至少部分核苷酸序列应是已知的

实现自动化。因此 PCR 不同于要花费数天时间完成的常规克隆，是只需要几个小时就能完成的“无细胞分子克隆”。这项技术如今普遍用于从基因组 DNA 或细胞中分离得到的 mRNA 开始的所研究基因的直接克隆（图 8-40）。

PCR 技术灵敏度非常高，甚至可以在样品中检出一个 DNA 分子。痕量的 RNA 可先经

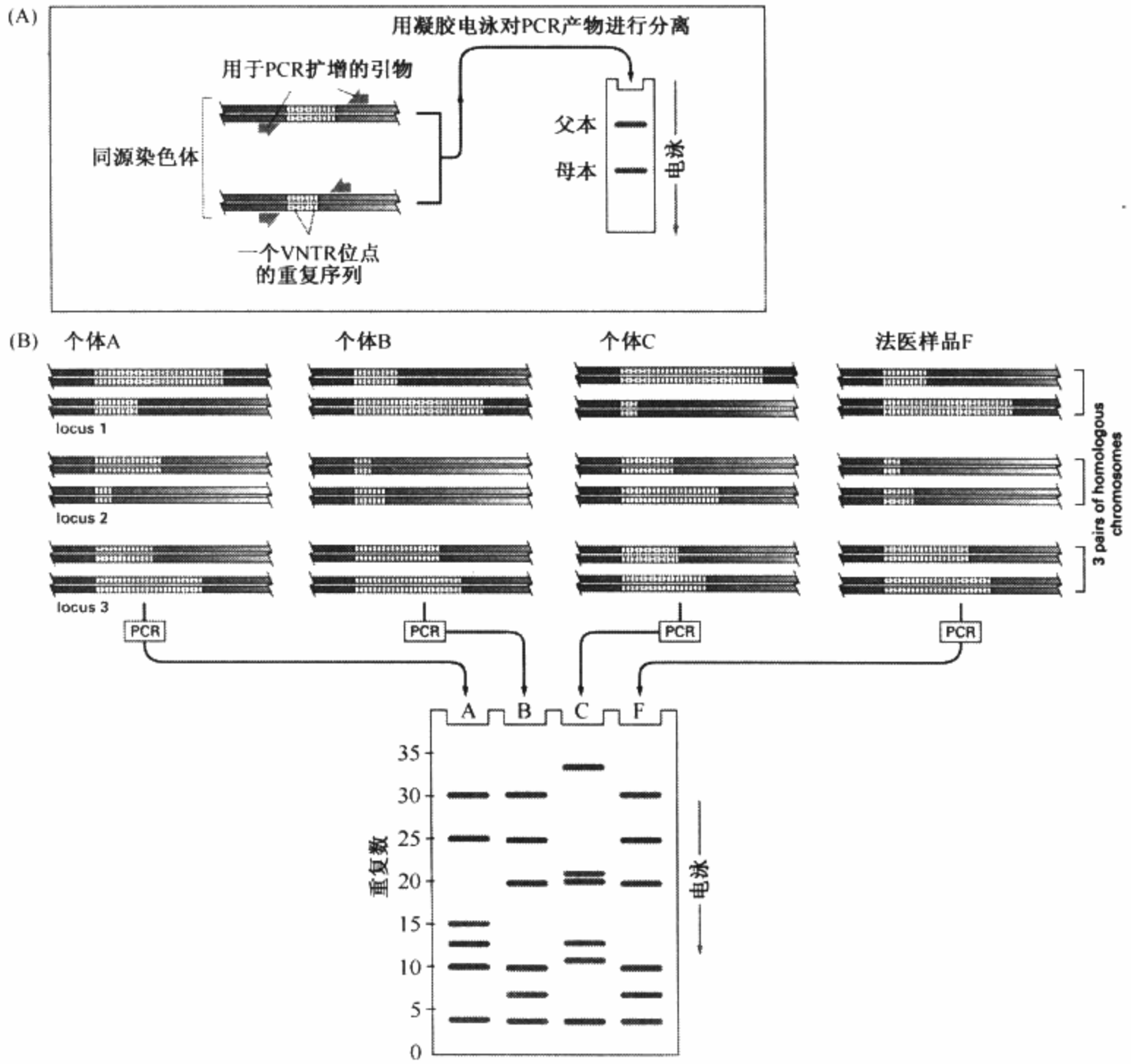


图 8-41　PCR 如何应用于法学

（A）在这个实验中产生差异的 DNA 序列中含有多个在人类基因组不同位点出现的短重复序列串，如 CACACA…。每个重复序列串的重复数在不同的个体中从 4～40 是高度多变的。一串这样的重复核酸序列叫做超变小卫星序列，即 VNTR（串联重复可变数）序列。这些序列在每个位点都具有可变性，个体通常从父亲和母亲那里分别继承一部分，而无亲缘关系的人之间通常不会含有相同的序列对。使用位于基因位点两端的引物进行 PCR 可以得到每个个体的两条 DNA 扩增片段条带，一条代表父本，一条代表母本。经过电泳，扩增所得的 DNA 的长度和产生的位置取决于该位点的具体重复数。（B）在上面的示意图中，对 3 个嫌疑人（个体 A、B、C）的 3 个相同的 VNTR 位点进行了分析（需要 3 对不同的特异性寡核苷酸引物），每个个体经聚丙烯酰胺凝胶电泳得到 6 条 DNA 条带。尽管这些个体有若干相同的条带，但总体的情况彼此不同。因此条带可以作为一个个体的近乎独特的“指纹”。在第 4 条泳道（F）中含有法医样品的反应产物。PCR 的起始产物可以是留在犯罪现场的一根头发或是少量血样。当鉴定 5～10 个 VNTR 位点时，两个个体具有相同基因模式的可能性只有大约 100 亿分之一。在这个例子中，个体 A 和 C 可以排除嫌疑，而个体 B 仍然是一个明显的嫌疑犯。如今类似的方法常常用于亲缘鉴定

反转录转录成 DNA 后用同样的方法分析。PCR 在基因疾病的诊断和少量病毒感染的检测上很大程度上取代了 RNA 印迹法。它在法医学上鉴定微量血迹或其他组织（甚至单个细胞）和通过基因“指纹”进行亲缘鉴定方面也有很大的发展空间（图 8-41）。

细胞内的蛋白质可以利用表达载体进行大量的表达

15 年以前，人们只能简单地研究相对丰富的细胞内蛋白质。从几百克的细胞中，采用连续的色谱分离步骤，可以纯化一种主要的细胞蛋白质（占细胞总蛋白质的 1%或以上），可以得到大约 0.1g（100mg）的纯蛋白质。这些蛋白质足以用于常规的氨基酸测序、生化反应的细节分析和生产抗体（可进一步用于蛋白质在细胞中的定位）。此外，如果可以长出可溶的晶体（一般比较困难），就可以利用 X 射线衍射技术测定这一蛋白质的三维结构，这一内容我们将在后面的章节中进行介绍。许多丰富蛋白质（包括血红蛋白、胰蛋白酶、免疫球蛋白和溶菌酶）的结构和功能，都是通过这种途径分析得到的。

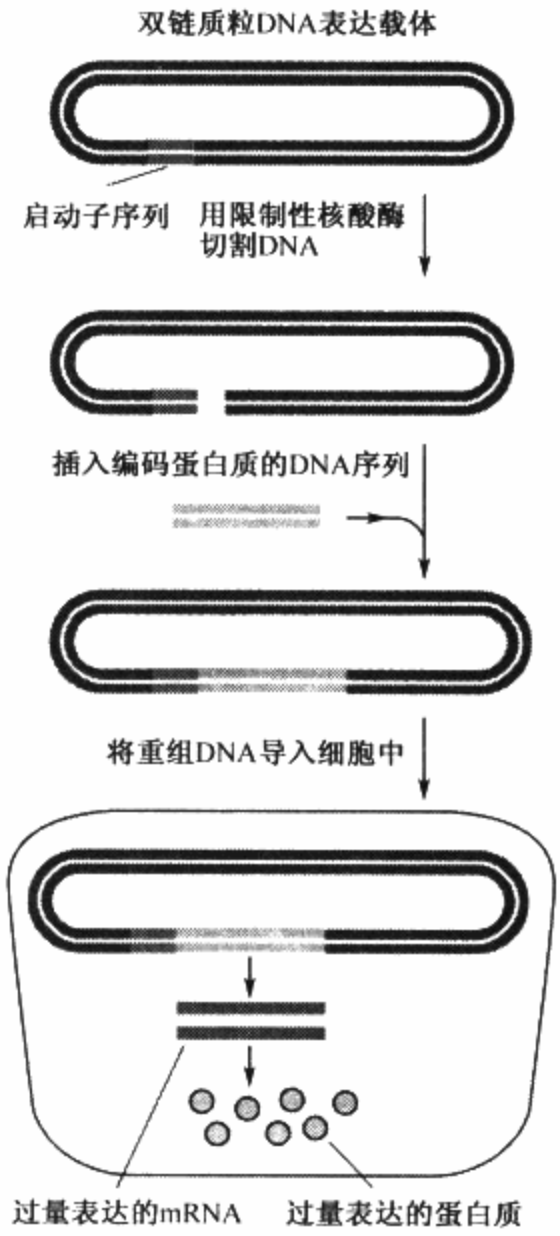


图 8-42　将编码蛋白的 DNA 序列克隆到表达载体中，并导入细胞中，产生大量的蛋白质

一种质粒载体经过改造后，含有一个活性很高的启动子，可以产生出对应于插入质粒的启动子邻近的蛋白质编码基因的大量的 mRNA。根据克隆载体的性质，可以将质粒导入细菌、酵母、昆虫或者哺乳细胞，使得插入的基因可以高效地转录并翻译成蛋白质

但是，在真核细胞的数以千计的不同蛋白质中，绝大多数是以极少量的方式存在的，包括那些具有至关重要功能的蛋白质。这些蛋白质大部分都很难（如果说不是不可能的）得到多于几个微克的纯品。使几乎不限量的生产任何细胞蛋白质成为可能，是 DNA 克隆技术和遗传工程对细胞生物学的最重要的贡献之一。

利用**表达载体**（图 8-42）可以在活细胞中生产大量的想要的蛋白质。这些表达载体是经过设计的普通载体，它们可以在转染的细菌、酵母、昆虫细胞和哺乳动物细胞中产生大量稳定的能有效翻译成蛋白质的信使 RNA。为了防止高水平表达的外源蛋白质影响转染的细胞的生长，表达载体常常通过设计，使外源信使 RNA 和蛋白质的合成可以延迟至收获细胞的前一刻（图 8-43）。

由于表达载体生产蛋白质是在细胞中进行的，在进行了细胞裂解之后，必须利用色谱方法把想要的蛋白质从宿主细胞所有的蛋白质中纯化出来；但由于在细胞裂解液中，这些蛋白质十分丰富（一般占细胞总蛋白的 1%～10%），通常几个步骤就可以完成纯化过程。许多表达载体设计添加了一个分子标签（一串组氨酸残基或者一个小的标记蛋白质），使表达的蛋白质可能采用简易的纯化过程，来进行亲和色谱层析，前面已经讨论过。现在有许多种经过改造的表达载体可用，每种都可以在表达蛋白质的细胞中发挥功能。通过这种方法，我们可以诱导来生产大量的有药用价值的蛋白质，如人胰岛素蛋白、生长激素、干扰素和接种用的病毒抗原。更一般地，这些方法使得大量生产所有蛋白质都成为可能，即使这些蛋白质在每个细胞中只有几个拷贝，生产所得的蛋白质可以用于细节

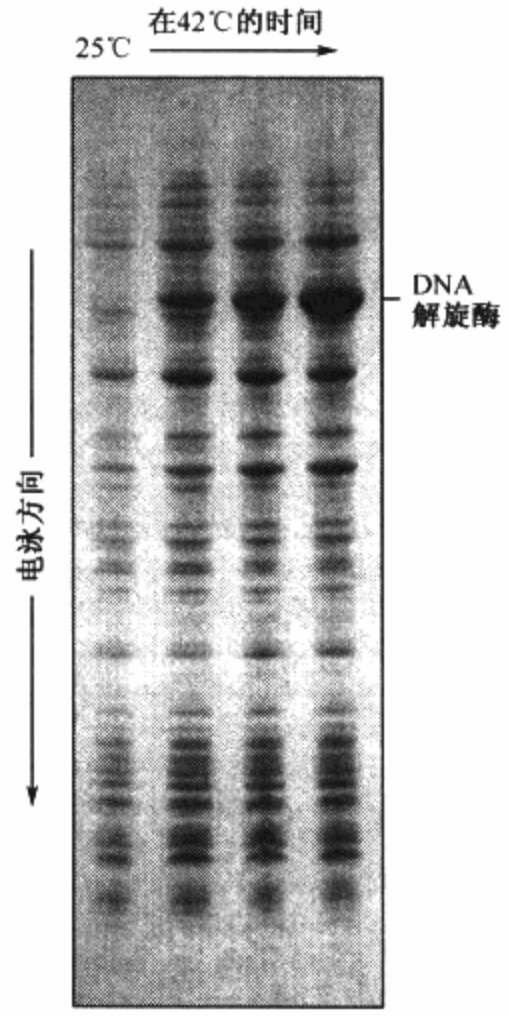


图 8-43　用质粒表达载体产生大量蛋白质

这个例子中，用 DNA 解旋酶的编码序列转染细菌细胞，这一序列的转录由一种只在 37℃及更高的温度下有活性的病毒启动子控制。细胞总蛋白质用 SDS -聚丙烯酰胺凝聚电泳进行分析，细胞样品中有在 25℃生长的细菌（不产生解旋酶蛋白）和放置于 42℃生长至 2h 的同样的细菌（解旋酶蛋白成为细胞裂解物中最丰富的蛋白质种类）（经 Jack Barry 许可）

性的结构和功能研究，这方面的内容我们将在后面的章节进行讨论（图 8-44）。

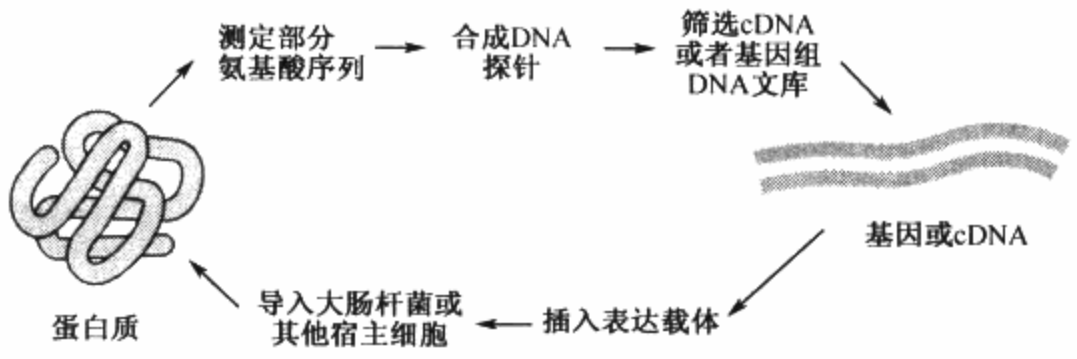


图 8-44　细胞分子生物学知识使在实验中从基因到蛋白质、从蛋白质到基因成为可能

小量纯蛋白质用于获得氨基酸序列的一部分，这就提供了序列信息，可以从 DNA 文库中克隆到相应的基因。一旦克隆到了基因，就可以将它的蛋白质编码序列插入到表达载体中，从遗传改造的细胞中产生大量的蛋白质

DNA 技术也可以用于大量生产任何已经分离出基因的 RNA 分子。纯 RNA 分子的应用，大大促进了 RNA 剪接、蛋白质合成、基于 RNA 的酶等的研究。大多数 RNA 痕量存在于细胞中，而且很难从细胞的其他组分（尤其是成千上万的其他 RNA 分子）中纯化出来。但对于任何感兴趣的 RNA 都可以通过高效的病毒 RNA 聚合酶转录 DNA 序列在体外进行有效的合成。合成的单一种类的 RNA 可以很容易地从 DNA 模板和 RNA 聚合酶中纯化出来。

小结

DNA 克隆可以将任何单拷贝的 DNA 或者 RNA 序列的特异部分从细胞中的数百万种序列中选择出来，并且可以以纯品的形式不限量的生产。将染色体 DNA 进行限制酶酶切，以及在自复制的遗传元件中插入 DNA 片段之后，就可以进行 DNA 序列的扩增。通常使用质粒载体进行这些操作，得到的基因组 DNA 文库收藏在上百万的细菌细胞中，每个保存一个不同的 DNA 克隆片段。这一文库中的可增殖的个体细胞可生产出大量的单一克隆的 DNA 片段。作为替代，聚合酶链反应可以利用一种纯的、热稳定的 DNA 聚合酶直接进行 DNA 克隆，前提是感兴趣的 DNA 序列已知。

用于获得与 mRNA 序列相对应的 DNA 克隆的流程和上述过程是一样的，只是要先得到 mRNA 的 DNA 拷贝即 cDNA。不同于基因组 DNA 克隆的是，cDNA 克隆缺乏内含子序列，使得它们成为分析基因的蛋白质产物的选择克隆。

核酸杂交反应提供了一种检测基因或者其他任何核苷酸序列的灵敏手段。在严格的杂交条件（仅仅一种完美的双螺旋能稳定存在的温度和溶剂组合）下，只有核苷酸序列可以近乎完美互补的两条链可以配对形成杂交螺旋。利用这种杂交反应的特异性，可以用标记了同位素或化学物质的任何核苷酸单链序列作为探针，在含有上百万种不同 DNA 和 RNA 序列的细胞或细胞提取物中找到互补链。这种探针广泛用于检测特异基因的核酸，以促进基因的纯化和鉴定，以及在细胞、组织、生物体中的定位。

纯化的 DNA 片段的核苷酸序列可利用基于双脱氧方法的高度自动化的技术进行测定。利用这一技术已经测定了成千上万个基因的完整 DNA 序列，并对许多生物体的基因组进行了完整的测序。不同生物体的基因组序列的比对，使我们可以追踪基因和生物体的进化关系，还可以用于发现新基因并预测其功能。

综上所述，这些技术使得从任何感兴趣的生物体中鉴定、分离、测定基因成为了可能。相关的技术使科学家可以大量生产出这些基因的蛋白质产物，用于结构和功能的分析和医药使用。

分析蛋白质的结构和功能

蛋白质执行生活细胞中的绝大多数任务。这种万能的大分子参与了细胞的各种进程：蛋白质复制和转录 DNA，生产、加工和分泌其他的蛋白质。它们调控细胞分裂、新陈代谢和细胞内外的物质和信息流动。理解细胞怎么运转需要先了解蛋白质如何行使功能。

生活细胞中的蛋白质都在做什么，这不是一个简单的问题。假定已经分离了一个未鉴定的蛋白质，它的结构和氨基酸序列分析表明它可以作为一个蛋白质激酶。简单的知道这个蛋白质可以在丝氨酸残基上加上一个磷酸基团，并不能揭示它怎样在生活细胞中行使功能。需要知道其他的信息才能理解利用它的生物化学活性进行的过程。这种激酶定位在细胞的哪里？它的蛋白质靶标是什么？它在哪种组织中有活性？它影响哪种通路？它在生物体的生长和发育中扮演什么角色？

在这一部分中，我们将讨论目前用于鉴定蛋白质结构和功能的方法。首先，我们要考察用于测定纯蛋白质的三维结构的方法。然后，我们讨论预测蛋白质功能的方法，这一方法是基于这一蛋白质与其他蛋白质相似性及其在细胞中的定位。最后，由于大多数蛋白质与其他蛋白质合作行动，我们介绍一些检测蛋白质-蛋白质相互作用的技术。但是，这些方法只能初

步说明一种蛋白质可能怎样在细胞中工作。在这一章的最后一部分，我们将讨论如何通过遗传学方法来研究和分析一种给定的蛋白质发挥功能的生物学过程。

蛋白质晶体的X射线衍射可以揭示蛋白质的准确结构

从蛋白质的氨基酸序列，人们一般可以预测出哪种二级结构元件（如跨膜α螺旋）会出现在蛋白质中。但是，目前还不可能从蛋白质的氨基酸序列推断其可靠的三维折叠结构，除非它的氨基酸序列与一个已知三维结构的蛋白质十分相似。在原子水平分辨率上显示分子（包括蛋白质）三维结构的主要方法是**X射线晶体学**。

X射线和光一样，也是一种形式的电磁辐射，但是它们有更短的波长，波长范围一般在0.1nm左右（相当于氢原子的直径）。如果一束窄的平行的X射线对准一个纯的蛋白质样品，大多数X射线会直接穿过它，但一小部分会被样品中的原子散射。如果这个样品是一个有序的晶体，那么散射波就会在某些点相互加强，当用适宜的检测仪器记录X射线时，会表现出多个衍射点（图8-45）。

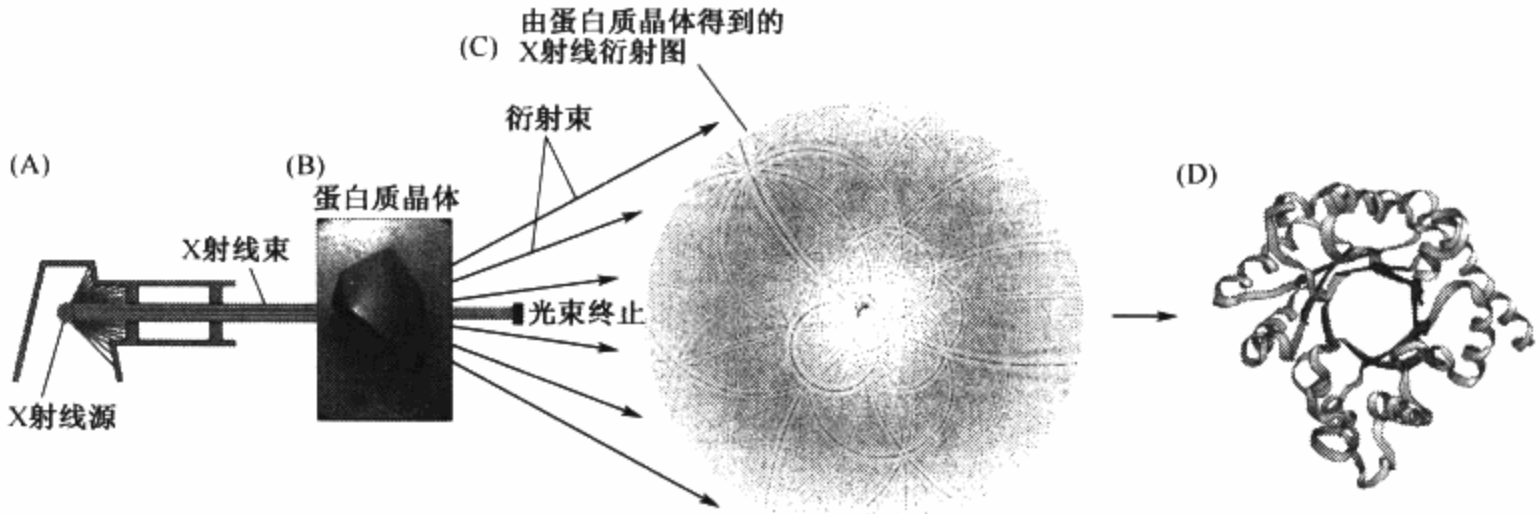


图8-45　X射线晶体学

（A）一束窄的平行的X射线对准一个有序排列的晶体。（B）这里展示的是一个二磷酸核酮糖羧化酶的蛋白质晶体，这种酶在光合磷酸化的二氧化碳固定过程中有核心作用。有些光束被晶体中的原子散射。散射波在某些点相互加强，表现为一种衍射点图案。（C）可利用这种衍射图，结合蛋白质的氨基酸序列，产生一个原子模型。（D）完整的原子模型难以通译，但是从X射线衍射数据得到的简单译本却可以清楚地展示蛋白质的结构特征（α螺旋，绿色；β链，红色）。注意：从（A）到（D）的图示中的组分没有按比例表示。［（B）经C. Branden许可；（C）经J. HajduI. Andersson许可；（D）摘自由B. Furugren提供的原始图片］

X射线衍射图中每个点的位置和亮度包含晶体中产生这个点的原子的位置信息。从晶体衍射图推断大分子的三维结构是一个复杂的工作，直到1960年对一个蛋白质分子的X射线结构测定才得以完成。但近年来X射线衍射分析逐渐变得自动化，现在最慢的一步可能要算适宜蛋白质晶体的产生。这一过程需要大量很纯的蛋白质，而且常常经过多年的反复试验，寻找合适的结晶条件。现在仍有许多蛋白质，特别是膜蛋白，还找不到合适的方法进行结晶。

衍射图的分析会产生一个复杂的三维结构电子密度图。解释这张图，即把图上的轮廓翻译成三维结构是一个复杂的过程，这一过程需要已知蛋白质的氨基酸序列。通过反复试验，用电脑可以给出序列和电子密度图之间最好、最可能的匹配。最终得到的原子模型的可靠性取决于最初的晶体学数据的分辨率：0.5nm分辨率可能产生多肽骨架的低分辨率图，而0.15nm分辨率却允许分子中所有的非氢原子都得到可靠的定位。

一个完整的原子模型一般太复杂了，难以直接进行通译，但是显示蛋白质核心结构特征的单一结构样式却可以容易地得到（专题 3-2）。现在大约有 10 000 种不同蛋白质的三维结构已经通过 X 射线晶体学或者核磁共振光谱学（见下面部分）进行了测定，足够形成各种具有共同结构的家族。这些结构或者蛋白质折叠，与它们的氨基酸序列相比，一般在进化上更加保守（图 3-15）。

X 射线晶体学技术也可以用于研究大分子复合物。最近这一方法已成功地解决了核糖体（一个由多种 RNA 和 50 多种蛋白质组成的大型复合细胞机器）的结构（图 6-64）。这一测定需要使用同步加速器，一种可以产生能分析这种大分子复合物晶体的 X 射线密度的辐射源。

分子结构也可以通过核磁共振光谱学进行测定

广泛使用**核磁共振（NMR）光谱学**分析小分子的结构已经有很多年了。现在，这种技术也逐渐用于研究小蛋白质或蛋白质结构域。不同于 X 射线晶体学，核磁共振不依赖于要有一个结晶的样品，它只需要将小体积的蛋白质浓溶液放置在强磁场中。

特定原子核，尤其是氢原子核，具有磁矩或者自旋：也就是说，它们像磁铁条一样有固有磁性。自旋沿着强磁场校准，但它也可以变成非校准的激发态，响应电磁辐射所采用射频脉冲。当激发态的氢原子回到校准态，它们发出射频辐射，可被测量并表示为一种频谱。发出辐射的性质决定于每个氢原子的环境，如果一个核处于激发态，邻近的其他核会影响其辐射的吸收和发射。因此，可以利用二维核磁共振技术来区分不同氨基酸残基中氢原子核的信号，判定和测量邻近氢原子核相互作用所产生的微小位移：位移的大小反映了一对相互作用的氢原子之间的距离。通过这种方法，核磁共振可以给出蛋白质分子各个部分之间的位置信息。用已知的氨基酸序列将这些位置信息组合起来，就可以在理论上计算蛋白质的三维结构（图 8-46）。

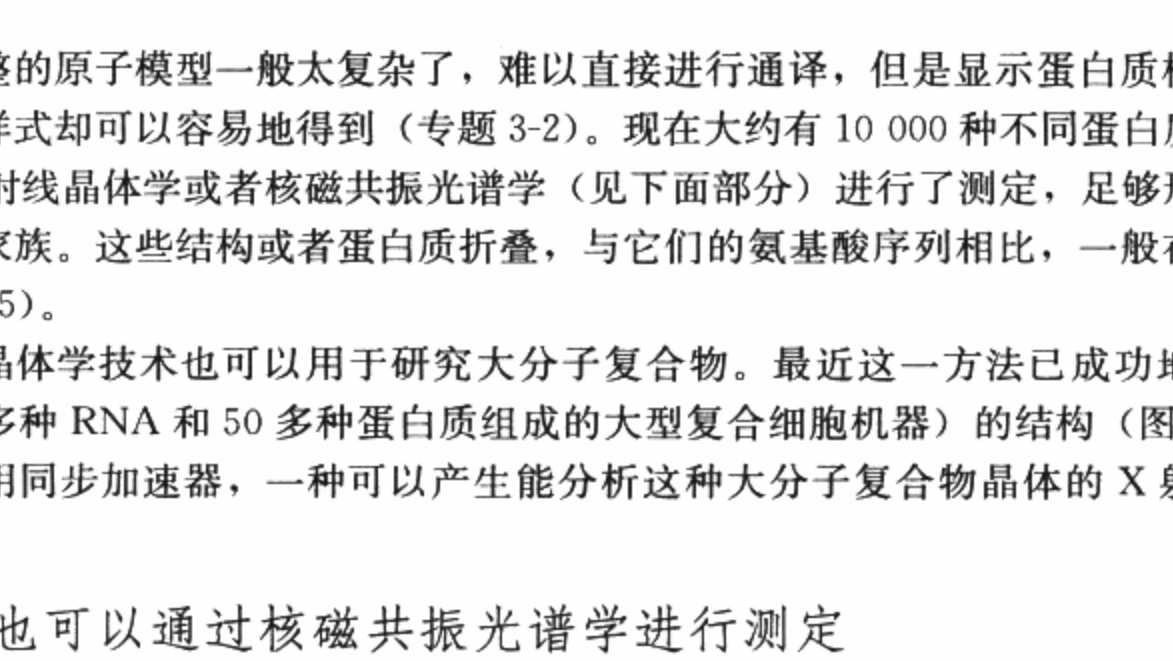


图 8-46　NMR 光谱

（A）NMR 仪器得到数据的示例。这个二维 NMR 光谱来源于纤维素酶的 C 端结构域。图中的点代表蛋白质中相邻氢原子的相互作用，反映了它们之间的距离。复杂的计算方法结合已知的氨基酸序列，可以得到可能的结构。（B）可以很好满足距离约束的 10 个结构，以互相重叠的方式表示，给出了可能的三维结构的良好推断（经 P. Kraulis 许可）

由于技术原因，分子质量在 20 000Da 或者更少的小蛋白质可用核磁共振光谱学测定结构。随着分子体积的增大，这一方法的分辨率逐渐丧失。但是新的技术进步让这种限制扩展到了 100 000Da，使大多数蛋白质可通过核磁共振进行结构分析。

当感兴趣的蛋白质不能结晶时，核磁共振方法就尤其有用了。由于核磁共振研究在溶液中进行，这种方法也提供了监测蛋白质结构变化的方便手段。例如，蛋白质折叠过程或者底物结合蛋白质的过程。核磁共振也广泛用于研究蛋白质之外的其他分子。例如，它也是测定 RNA 分子和复杂的糖蛋白糖侧链等结构的有用方法。

一些在 X 射线晶体学和核磁共振发展过程中的里程碑事件列在表 8-8 中。

表 8-8　X 射线晶体学和 NMR 发展中的里程碑和它们在生物大分子中的应用

1864 年	Hoppe-Seyler 结晶并命名了血红蛋白
1895 年	Röntgen 观察到当阴极射线（电子）撞击金属靶时，会产生贯穿辐射，并将其命名为 X 射线
1912 年	Von Laue 用 X 射线穿过硫酸锌晶体，得到了第一张 X 射线衍射图 W. L. Bragg 提出了 X 射线衍射图和产生此图的晶体中原子排列之间的简单关系
1926 年	Summer 从刀豆提取物中得到了尿素酶晶体，并证明这些蛋白质具有催化活性
1931 年	Pauling 发表了关于“化学键的本质”的第一篇论文，详述了共价键形成的规律
1934 年	Bernal 和 Crowfoot 发表了从胃蛋白酶晶体得到的第一张详细的 X 射线衍射图
1935 年	Patterson 发明了从 X 射线数据测定原子间距的分析方法
1941 年	Astbury 得到了首张 DNA 的 X 射线衍射图
1951 年	Pauling 和 Corey 提出了一条 L-氨基酸链的螺旋构象结构（α 螺旋）和 β 折叠结构，之后在许多蛋白质中发现了这两种结构
1953 年	Watson 和 Crick 提出了 DNA 的双螺旋模型，这一模型基于 Franklin 和 Wilkins 得到的 X 射线衍射图
1954 年	Perutz 和同事发展出了重原子方法来解决蛋白质晶体学中的相位问题
1960 年	Kendrew 描述了第一个蛋白质（抹香鲸肌红蛋白）详细结构，分辨率为 0.2nm；Perutz 提出了较大的血红蛋白的较低分辨率的结构
1966 年	Phillips 描述了溶菌酶的结构，这是第一个详细分析结构的酶
1971 年	Jeener 建议应用二维核磁共振；Wuthrich 和同事在 20 世纪 80 年代初最先利用这一方法解决了一个蛋白质的结构
1977～1978 年	Holmes 和 Klug 测定了烟草花叶病毒的结构；Harrison 和 Rossman 测定了两种小的球形病毒的结构
1985 年	Michel、Deisenhofer 和同事利用 X 射线晶体学测定了第一个跨膜蛋白（一个细菌反应中心）的结构。1975～1990 年，Henderson 和同事通过高分辨率的电子显微镜得到了细菌视紫红质（一种跨膜蛋白）的结构

序列相似性可以提供蛋白质功能的线索

由于基因组数据库中的蛋白质和核酸序列不断增加，基因和它所编码的蛋白质的功能一般可以通过与以前鉴定出的基因进行简单的序列对比来进行预测。因为氨基酸序列决定了蛋白质的结构，结构可以指示生物化学功能，所以具有相似氨基酸序列的蛋白质通常具有相似的生物化学功能，即使这些蛋白质是在关系很远的生物体中发现的。因此目前测定一个新发现的蛋白质的功能，通常要先搜索和它具有氨基酸序列相似性的已经鉴定出的蛋白质。

在一系列的已知序列中搜索同源基因或蛋白质，要在因特网上进行，只要选择一个数据库并且输入需要的序列即可。序列比对的程序（最流行的是 BLAST 和 FASTA）扫描数据库

寻找相似序列，使提交的序列在存档的序列上滑动直至一串残基遇到完全或者部分的对应序列（图 8-47）。无论在核苷酸还是蛋白质序列上进行，一个复杂的搜索也只要几分钟就可以得到结果。这种比对可以用来预测单个蛋白质、蛋白质家族，甚至一个刚刚测序的生物体的所有蛋白质的功能。

```
 Score =  399 bits (1025), Expect = e-111
 Identities = 198/290 (68%), Positives = 241/290 (82%), Gaps = 1/290

Query: 57  MENFQKVEKIGEGTYGVVYKARNKLTGEVVALKKIRLDTETEGVPSTAIREISLLKELNH 116
           ME ++KVEKIGEGTYGVVYKA +K T E +ALKKIRL+ E EGVPSTAIREISLLKE+NH
Sbjct: 1   MEQYEKVEKIGEGTYGVVYKALDKATNETIALKKIRLEQEDEGVPSTAIREISLLKEMNH 60

Query: 117 PNIVKLLDVIHTENKLYLVFEFLHQDLKKFMDASALTGIPLPLIKSYLFQLLQGLAFCHS 176
            NIV+L DV+H+E ++YLVFE+L  DLKKFMD+          LIKSYL+Q+L G+A+CHS
Sbjct: 61  GNIVRLHDVVHSEKRIYLVFEYLDLDLKKFMDSCPEFAKNPTLIKSYLYQILHGVAYCHS 120

Query: 177 HRVLHRDLKPQNLLINTE-GAIKLADFGLARAFGVPVRTYTHEVVTLWYRAPEILLGCKY 235
           HRVLHRDLKPQNLLI+    A+KLADFGLARAFG+PVRT+THEVVTLWYRAPEILLG +
Sbjct: 121 HRVLHRDLKPQNLLIDRRTNALKLADFGLARAFGIPVRTFTHEVVTLWYRAPEILLGARQ 180

Query: 236 YSTAVDIWSLGCIFAEMVTRRALFPGDSEIDQLFRIFRTLGTPDEVVWPGVTSMPDYKPS 295
           YST VD+WS+GCIFAEMV ++ LFPGDSEID+LF+IFR LGTP+E  WPGV+ +PD+K +
Sbjct: 181 YSTPVDVWSVGCIFAEMVNQKPLFPGDSEIDELFKIFRILGTPNEQSWPGVSCLPDFKTA 240

Query: 296 FPKWARQDFSKVVPPLDEDGRSLLSQMLHYDPNKRISAKAALAHPFFQDV 345
           FP+W  QD + VVP LD  G  LLS+ML Y+P+KRI+A+ AL H +F+D+
Sbjct: 241 FPRWQAQDLATVVPNLDPAGLDLLSKMLRYEPSKRITARQALEHEYFKDL 290
```

图 8-47　BLAST 搜索的结果

可以通过搜索序列数据库来找到相似的氨基酸或者核苷酸序列。这里通过搜索，找到与人类细胞周期调节蛋白 cdc2（Query，查询）相似的蛋白质是玉米 cdc2（Subject，目标），它与人类的 cdc 在氨基酸序列上有 68%的一致性和 82%的相似性。这一比对开始于查询蛋白质的第 57 个残基，说明人类蛋白质有一段 N 端区域是玉米蛋白质所没有的。绿色方块指示序列的不同性，黄色长条总结了相似性：当两段氨基酸序列一致时，残基就显示出来；保守的氨基酸替代用加号（+）表示。为了使两个序列得到最大的匹配，只引入了一个小的缺口，用红色的箭头在查询序列的第 194 位指出。比对得分（Score，得分）考虑了替代和缺口的罚分；比对得分越高，序列匹配得越好。比对的意义反映在期望值（E）上，它代表了等同于或者好于偶然事件期望值得分的比对数量。E 值越低，比对的意义越大；非常低的 E 值具有特定的意义。E 值大大高于 0.1 不太可能反映真实的相关性

但是，通过序列分析进行的预测一般只能作为指导进一步实验研究的工具。

融合蛋白可以用于分析蛋白质功能和在活细胞中追踪蛋白质

蛋白质在细胞中的定位一般可以给出一些它的功能信息。例如，当细胞遇到生长因子时从细胞质进入细胞核中的蛋白质，可能在这种因子相应的基因表达调控中有作用。蛋白质一般含有一段决定其细胞定位的短的氨基酸序列。例如，大多数细胞核蛋白含有一个或者多个特殊的氨基酸短序列，这种序列可以作为它们在细胞质中合成后运入细胞核的信号（讨论见第 12 章）。蛋白质的这些特殊区域可以通过下面的方法进行鉴定：将蛋白质与一种缺少这种特殊区域的且易检测的蛋白质进行融合，然后追踪这种替代蛋白质在细胞中的行为。这种融合蛋白可以通过前面讨论的 DNA 重组技术产生。

另一种既能在细胞中跟踪蛋白质又能快速纯化蛋白质的常用策略是抗原决定簇标记。在这种方法中，产生了一种含有待分析的完整蛋白质和一段 8～12 个氨基酸的短肽（一个抗原决定簇）的融合蛋白，其上的短肽可被商品化的抗体识别。接下来，可以利用抗抗原决定簇的抗体和带有标记的二抗（可被光学显微镜或电子显微镜监测），特异地检测这种融合蛋白，

即使有正常蛋白质过量存在时也可行（图 8-48）。

现在，大量的蛋白质通过一种叫做**绿色荧光蛋白（GFP）**的荧光标记物在活细胞内进行示踪。将 GFP 标记在蛋白质上，只需将 GFP 的基因接到编码所需蛋白质的基因的一端。在大多数情况下，得到的 GFP 融合蛋白和原来的蛋白质有相同的表现，而且可以通过荧光显微镜监测它在细胞中的移动。GFP 融合蛋白策略已经成为一种测定任何感兴趣的蛋白质在细胞中的分布和运动的标准方法。

GFP 及其不同颜色的衍生物，也可以用于监测蛋白质-蛋白质相互作用。两种感兴趣的蛋白质分别被标记上不同的荧光染料，一种荧光染料的发射光谱与第二种荧光染料的吸收光谱互相叠加。如果这两种蛋白质（结合有荧光染料）相互靠近，达到 1～10nm，吸收光的能量就会从一种荧光染料传到另一种。这种能量传递，叫做**荧光共振能量传递（FRET）**，可以通过照射第一种荧光染料和测量第二种荧光染料的发射光谱来进行测定（图 8-49）。通过使用 GFP 的不同光谱的变异物作为荧光染料，人们可以监测细胞中任何两种蛋白质之间的相互作用。

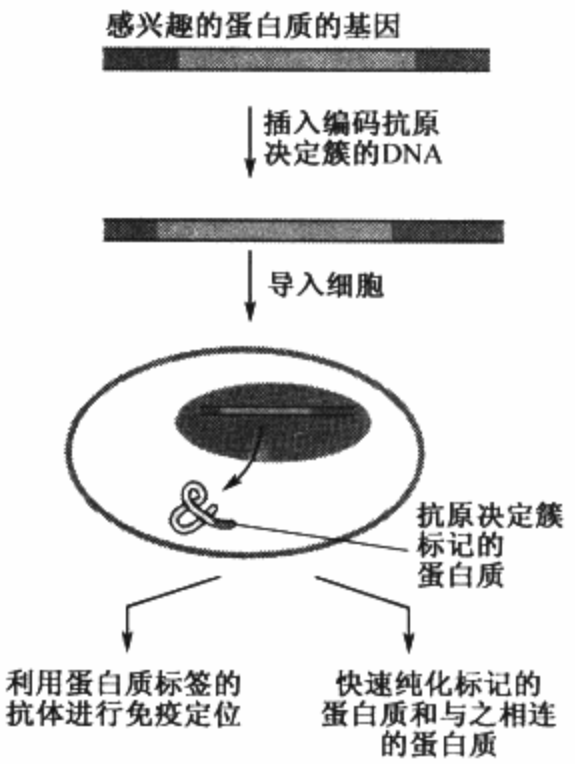


图 8-48　抗原决定簇标记可以用于蛋白质的定位和纯化

利用标准的遗传工程技术，一个短的抗原决定簇标签可以添加到感兴趣的蛋白质上。得到的蛋白质包括一段可被商品化的抗体识别的短肽。标记的抗体可用于跟踪蛋白质在细胞中的定位或者通过免疫沉淀或亲和层析来进行纯化

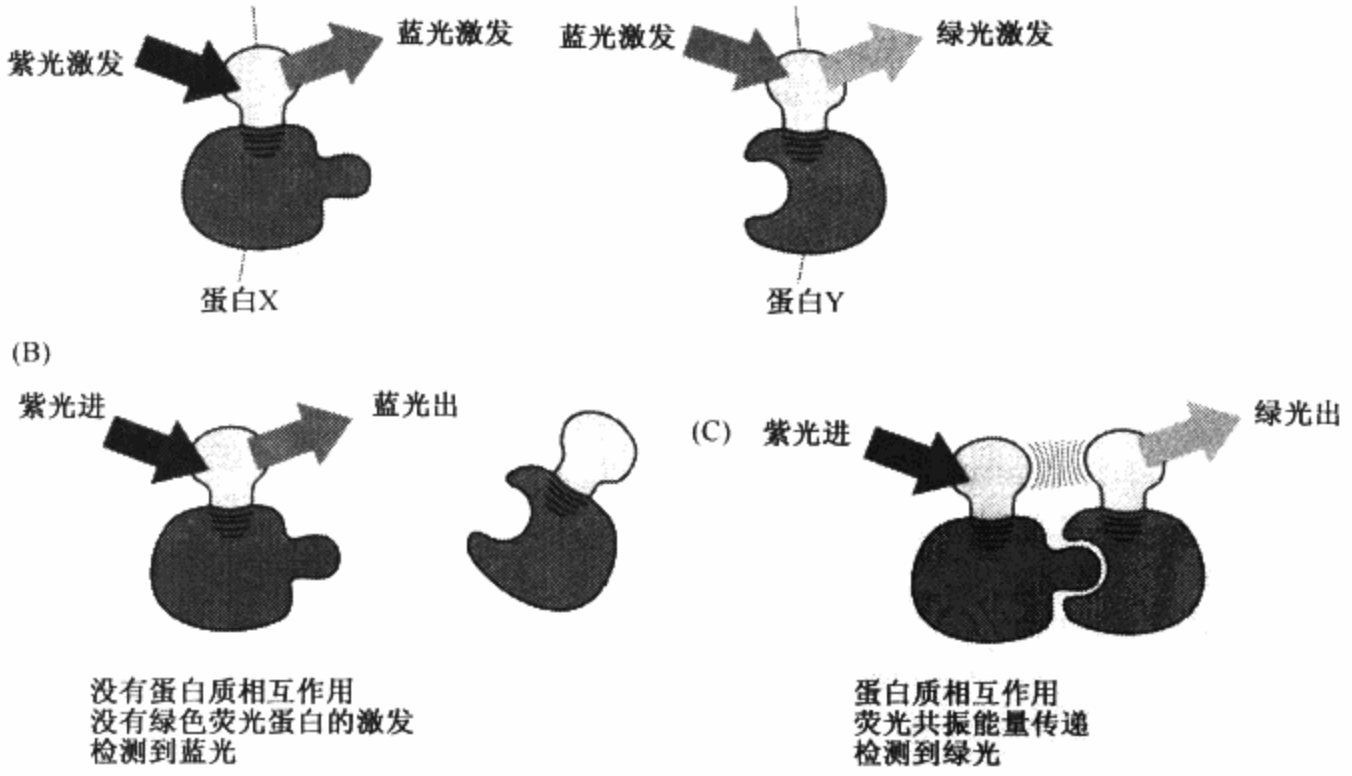


图 8-49　荧光共振能量传递（FRET）

为了测定两种蛋白质在细胞中是否发生相互作用（或者什么时候相互作用），首先要产生含有 GFP 不同衍生物的融合蛋白。（A）在这个例子中，蛋白 X 与蓝色荧光蛋白相偶联，可以被紫光（370～440nm）激发，并发出蓝光（440～480nm）；蛋白 Y 与绿色荧光蛋白相偶联，可以被蓝光激发，并发出绿光（510nm）。（B）如果蛋白 X 和 Y 不发生相互作用，用紫光照射样品只能产生来自蓝色荧光蛋白的荧光。（C）当蛋白 X 和 Y 发生相互作用时，荧光共振能量传递就发生了。用紫光照射样品会激发蓝色荧光蛋白，发出的蓝色荧光会接着激发绿色荧光蛋白，最终产生绿色荧光。荧光色素必须非常近，距离在 1～10nm，才能产生荧光共振能量传递。由于并非每一分子的蛋白 X 和蛋白 Y 都是一直结合在一起的，因此仍可以检测一些蓝光。但是当两种蛋白质开始相互作用之后，发射者 GFP 发光量下降，同时接受者 GFP 发光量上升

亲和层析和免疫沉淀可以测定相互关联的蛋白质

由于细胞中的大多数蛋白质以复合物的一部分和其他蛋白质一起发挥功能，鉴定它们的结合物成为一种初步鉴别其生物学作用的重要方法。如果一个未鉴定的蛋白质与细胞内一个已知功能的蛋白质结合，那么它的功能也与此相关。例如，如果发现一个蛋白质是蛋白酶体复合物的一部分，它可能与降解损坏的或者错误折叠的蛋白质有关。

蛋白质亲和层析是一种用于分离和鉴定有相互作用的蛋白质的方法。为了捕获相互作用的蛋白质，靶蛋白结合到填充柱子的多聚物珠子上，细胞蛋白质流过柱子，与靶蛋白相互作用的蛋白质就会黏附到亲和基质上［图 8-11（C）］。然后这些蛋白质被洗脱下来，通过质谱或者其他合适的方法进行鉴定。

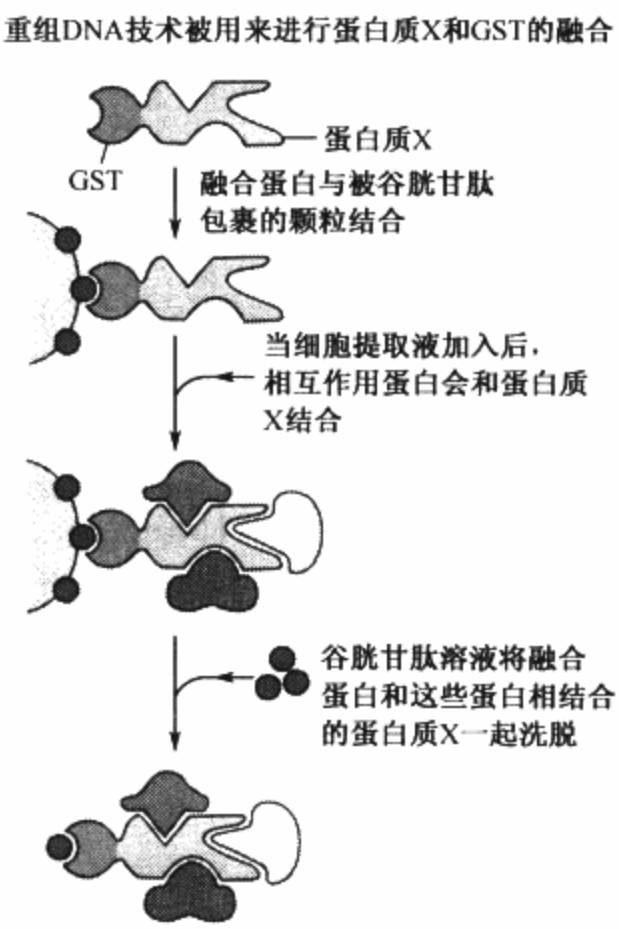


图 8-50　利用 GST 标记的融合蛋白纯化蛋白质复合物

GST 融合蛋白，可以通过标准的 DNA 重组技术产生，可以被由谷胱甘肽包裹的颗粒组成的亲和层析柱捕获。为了找到和蛋白 X 结合的蛋白质，细胞提取液需要通过这个柱子。然后，杂交蛋白和它的结合伙伴可以用谷胱甘肽洗脱下来。这些相互作用的蛋白质的身份可以通过质谱鉴定（图 8-20）。在替代方法中，从产生 GST 融合蛋白的细胞得到细胞提取物后，直接通过谷胱甘肽亲和层析柱。GST 融合蛋白和细胞中与它结合的蛋白质一起被保留在柱子上。亲和层析柱也可以用 GST 抗体或者其他易得到的小蛋白质或抗原决定簇来制备（图 8-48）

鉴定相互结合紧密的蛋白质的最简单方法可能是**免疫共沉淀**。这种方法，利用一种抗体识别一种特殊的靶蛋白；与抗体结合的亲和物质与固体基质偶联，然后将复合物从溶液中拖到试管的底部。如果这种蛋白质在被抗体捕获的过程中与另一种蛋白质紧密相连，它的伙伴也会同时沉淀。这种方法可用于鉴定细胞内复合物的组分蛋白质，包括那些只发生暂时作用的，如细胞被信号通路激活时（讨论见第 15 章）。

免疫共沉淀技术需要得到对已知的细胞内的靶蛋白具有高度特异性的抗体，而这些抗体不是总可以得到的。克服这种困难的方法之一是利用重组 DNA 技术，添加一个抗原决定簇标签（图 8-48）或者将靶蛋白与一个已经鉴定的标记蛋白融合，如谷胱甘肽 S 转移酶（GST）。然后，用直接抗抗原决定簇标签或者标记蛋白质的商业化的抗体，来沉淀整个融合蛋白，包括与感兴趣的蛋白质相关联的任何细胞内的蛋白质。如果蛋白质是与 GST 融合的，也许根本不需要抗体：杂交物和它的结合物可以直接用包裹了谷胱甘肽的珠子进行选择（图 8-50）。

除了在柱子上或者在试管中捕获蛋白质复合物，研究者还发展了高密度蛋白质芯片来研究蛋白质功能和蛋白质相互作用。这些芯片含有点在玻璃板上或者固定在小池子里的成千上万种蛋白质和抗体，可以一次检测大量蛋白质的生化活性和结合特性。用这种芯片检测蛋白质相互作用，需要将标记的蛋白质和每一种固定在芯片板上的靶蛋白温育，然后检测标记的蛋白质和哪些蛋白质结合。

蛋白质-蛋白质相互作用可以用双杂交系统进行鉴定

免疫共沉淀和亲和层析等方法可用于相互作用蛋白质的物理分离。成功的分离只产生一种可以被质谱鉴定的蛋白质，在进一步鉴定它的活性（蛋白质-蛋白质相互作用的实质）之前要先得到并克隆出它的基因。

其他的技术可以同时分离出相互作用的蛋白质和编码它们的基因。首先我们要讨论的方法是**双杂交系统**，利用一个报道基因检测酵母细胞核中一对蛋白质的物理相互作用。设计这一系统，使得当靶蛋白在细胞中结合另一蛋白质时，它们的相互作用将一个转录激活子的两半带到一起，就可以开启报道基因的表达。

这种技术利用了基因激活子蛋白质的模块性质（图 7-42）。这些蛋白质既可以和 DNA 结合，又可以激活转录，这两种活动通常由两个独立的结构域完成。利用重组 DNA 技术，编码靶蛋白的 DNA 序列与编码基因激活子的 DNA 结合结构域的 DNA 融合。当这一重组物进入酵母，细胞就会产生靶蛋白连着 DNA 结合结构域的重组蛋白（图 8-51）。这种蛋白质与报道基因的调节区域结合，作为“饵”，来捕获酵母细胞中与靶蛋白相互作用的蛋白质。为了得到一系列可能的结合伙伴，编码基因激活子激活结构域的 DNA 要与从 cDNA 文库中得到的大量 DNA 片段进行连接。这些收集的基因片段，即“捕获物”，一个一个的引入含有饵的酵母细胞。如果酵母细胞得到了一个表达饵蛋白结合伙伴的 DNA 克隆，转录激活子的两半就会联合，开启报道基因（图 8-51）。筛选表达报道基因的细胞，就可以得到编码捕获蛋白的基因或者基因片段，并通过核苷酸测序来进行鉴定。

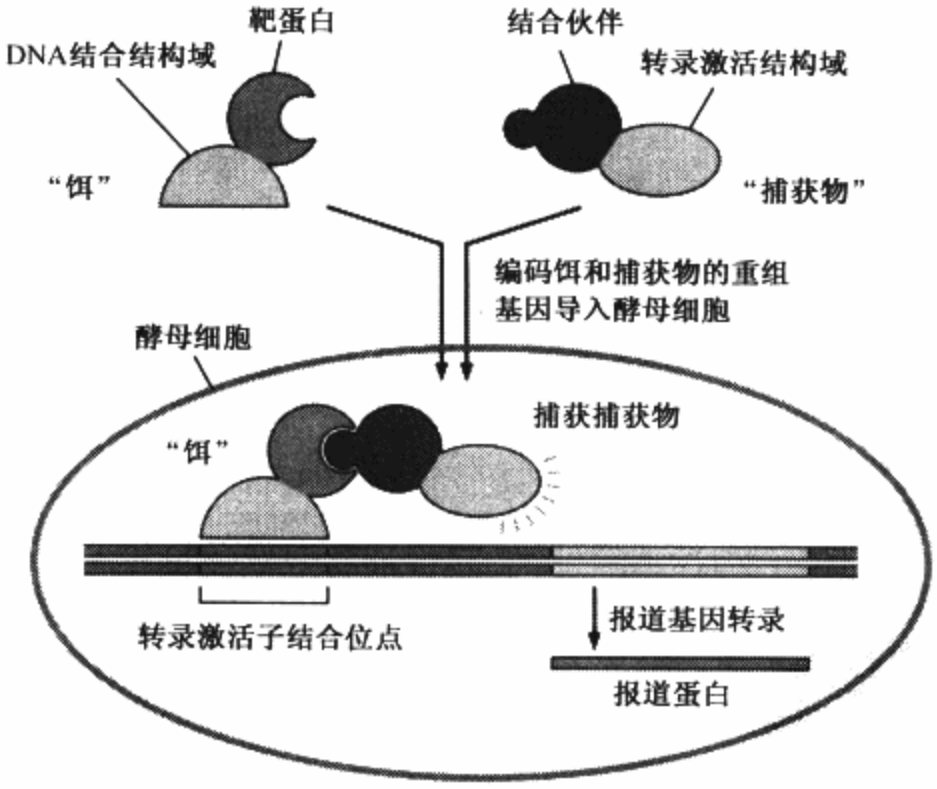


图 8-51　检测蛋白质-蛋白质相互作用的酵母双杂交系统

靶蛋白与 DNA 结合结构域融合，使它作为“饵”定位在报道基因的调节区域。当这种靶蛋白与另一个特别设计的细胞核内的蛋白质（“捕获物”）结合时，它们的相互作用将一个转录激活子的两半带到一起，就可以开启报道基因的表达。报道基因通常使细胞在选择性培养基上可以生长。饵和捕获物融合蛋白通过标准的 DNA 重组技术产生。在大多数情况下，一个单独的饵蛋白用来在大量的捕获蛋白质中捕获相互作用伙伴，这种捕获蛋白质是通过将编码转录激活子的激活结构域的 DNA 连接到 cDNA 文库的大量 DNA 片段上产生的

尽管听起来很复杂，在实验室中应用双杂交系统还是相对简单的。尽管蛋白质-蛋白质相互作用发生在酵母细胞核中，但来自细胞每个部分和来自任何生物体的蛋白质都可以通过这

一方法进行研究。在酵母中研究过的数千种蛋白质-蛋白质相互作用中，一半都是通过双杂交筛选发现的。

可以按比例增强双杂交系统，来对一种生物体产生的所有蛋白质间发生的相互作用进行作图。在这种情况下，针对每个细胞内蛋白质要生产一系列饵融合蛋白，每种饵融合蛋白要进入一个独立的酵母细胞。然后，这些细胞要与含有捕获蛋白文库的酵母进行繁殖，接着那些蛋白质-蛋白质相互作用呈阳性的细胞被鉴定出来。通过这种方法，一个关于酵母中 6000 种蛋白质的关联图就产生了（图 3-78），针对线虫和果蝇的蛋白质相互作用的相似计划正在进行中。

一种叫做反向双杂交系统的相关技术，可以用于鉴别能破坏蛋白质-蛋白质相互作用的突变体或者化学复合物。这种方法中，报道基因可以用另一种基因来代替，这种基因在饵蛋白和捕获蛋白发生相互作用时会杀死细胞。因此，只有由于突变或者复合物阻断了蛋白质的相互结合的细胞可以存活下来。类似于基因敲除（之后我们会讨论到），消除一种特定的分子间相互作用可以揭示细胞中参与这一作用的蛋白质的功能。此外，有选择地阻断蛋白质相互作用的复合物有药用价值。例如，可以阻断病毒与人细胞表面的受体蛋白结合的药物可以帮助人们避免病毒感染。

噬菌体展示的方法也可以检测蛋白质相互作用

另一种检测蛋白质-蛋白质相互作用的方法是将基因导入可以浸染大肠杆菌的病毒（噬菌体）中。在这种方法中，编码感兴趣的蛋白质或者此蛋白质的一小段肽段的基因，与一个编码病毒衣壳蛋白的基因融合。当这种病毒浸染大肠杆菌时，它会复制并产生在衣壳上展示杂交蛋白质的病毒颗粒［图 8-52（A)］。接下来，这种噬菌体可用于在很大的潜在靶蛋白的库中捕捉结合伙伴。

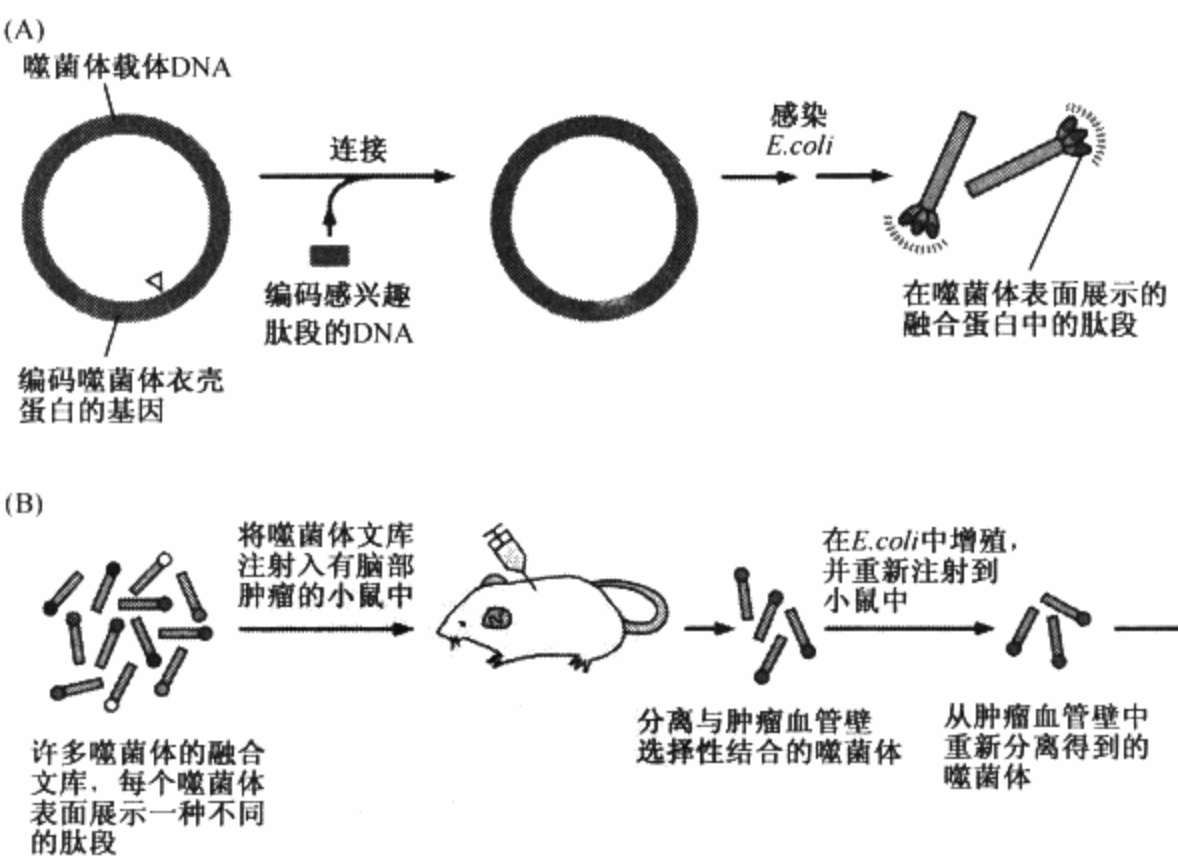


图 8-52　用于研究蛋白相互作用的噬菌体展示方法

（A）制备噬菌体。编码所需肽段的 DNA 与噬菌体载体进行连接，与编码病毒衣壳蛋白的基因融合。改造的噬菌体导入大肠杆菌，产生展示杂交衣壳蛋白的噬菌体，杂交的衣壳蛋白含有所需肽段。（B）产生含有无数种肽段的噬菌体文库。在这个例子将噬菌体文库注射入有脑部肿瘤的小鼠中，筛选与肿瘤血管内壁选择性结合的噬菌体，分离并使其增殖。然后，就可以从纯化的噬菌体中分离出与肿瘤血管壁特异结合的肽段。这种肽段可用于确定肿瘤的药物或者毒素

但是，这种**噬菌体展示**的方法需要针对选定的靶标来筛选大量的蛋白质或者多肽。这一过程首先需要产生一个融合蛋白质文库，类似于双杂交系统的捕获蛋白质文库。然后，筛选这种文库中的噬菌体，寻找与感兴趣的纯蛋白质结合的噬菌体。例如，可以将噬菌体文库通过一个含有固定靶蛋白的亲和层析柱。展示可与靶蛋白紧密结合的蛋白质或者多肽的病毒，会被捕获到柱子上，并被大量的靶蛋白洗脱下来。收集这些含有编码相互作用蛋白质或多肽的 DNA 片段的噬菌体，并让它们在大肠杆菌中进行复制。然后，就可以重新获得每个噬菌体中的 DNA，并对它们进行序列测定，来确定与靶蛋白结合的蛋白质或者多肽伙伴。类似的方法已经用于分离与人类肿瘤相关的和血管内壁特异结合的多肽。这些多肽目前正在进行试剂测试，来生产可直接作用于那些肿瘤的治疗性抗癌复合物［图 8-52(B)］。

噬菌体展示还用于产生可以识别特异靶分子或细胞的单克隆抗体。在这种情况下，要筛选一个表达抗体分子合适部分的噬菌体文库，来获得与靶抗原结合的噬菌体。

蛋白质相互作用可以利用表面等离子体共振来进行实时监测

一旦已知两种蛋白质或者一种蛋白质和一种小分子可以互相结合，更加详细的描述它们的相互作用就变得很重要了。蛋白质可以永久地相互结合，或者只是短暂遭遇后暂时结合在一起，这些动态的相互作用一般是通过可逆的修饰（如磷酸化）、配基结合或者其他竞争同一个结合位点的蛋白质存在或缺乏来调节的。

着手研究这些复杂的过程时，必须首先测定两种蛋白质结合的紧密程度、分子复合物结合和分离的快慢程度、外界因素如何影响这些参数。在本章中，有许多不同的技术可以用来研究蛋白质-蛋白质相互作用，每种都有自己的优点和缺点。一种监测蛋白质结合动态过程的特别有用的方法是**表面等离子体共振（SPR）**。利用 SPR 方法已经鉴定了很多种分子间相互作用，包括抗体-抗原结合，配基-受体偶合，以及蛋白质与 DNA、糖、小分子和其他蛋白质的结合。

SPR 通过监测可能的结合分子水溶液和带有固定的饵蛋白的生物传感器间界面上光束的反射来检测结合作用。饵蛋白结合到玻璃棱镜面上一层非常薄的金属涂层上（图 8-53）。一束光以特定的角度（叫做共振角）通过棱镜，一部分光能与金属薄膜的电子云相互作用，产生等离子体，电子以合适的角度在薄膜平面上振动，在上下表面间来回弹跳，像砝码在弹簧上一样。接着，等离子体在金属表面上下产生了一个短距离的电场，大概有光的波长那样长。电场范围内的环境中任何的组分改变都会引起共振角的可测量的变化。

为了测量结合程度，让含有可能与固定的饵蛋白发生相互作用的蛋白质或分子的溶液流过生物传感器表面。当蛋白质与饵结合时，金属表面的分子复合物的组分发生改变，引起共振角的变化（图 8-53）。实时监控共振角的变化，这种变化可以反映分子和饵蛋白结合和分离的动态过程。当分子发生相互作用时测量结合速率（k_{on}），当用缓冲液从传感器表面将结合分子洗脱下来时测量分离速度（k_{off}）。用 k_{off} 除以 k_{on} 来计算结合常数（K）。除了测定动力学过程，SPR 也可以用于测定每个复合物中结合的分子数目：SPR 信号改变的大小与固定的复合物的质量呈正比。

SPR 方法特别实用，因为它只需要少量的蛋白质，蛋白质不需要任何的标记，而且蛋白质-蛋白质相互作用可以进行实时监控。

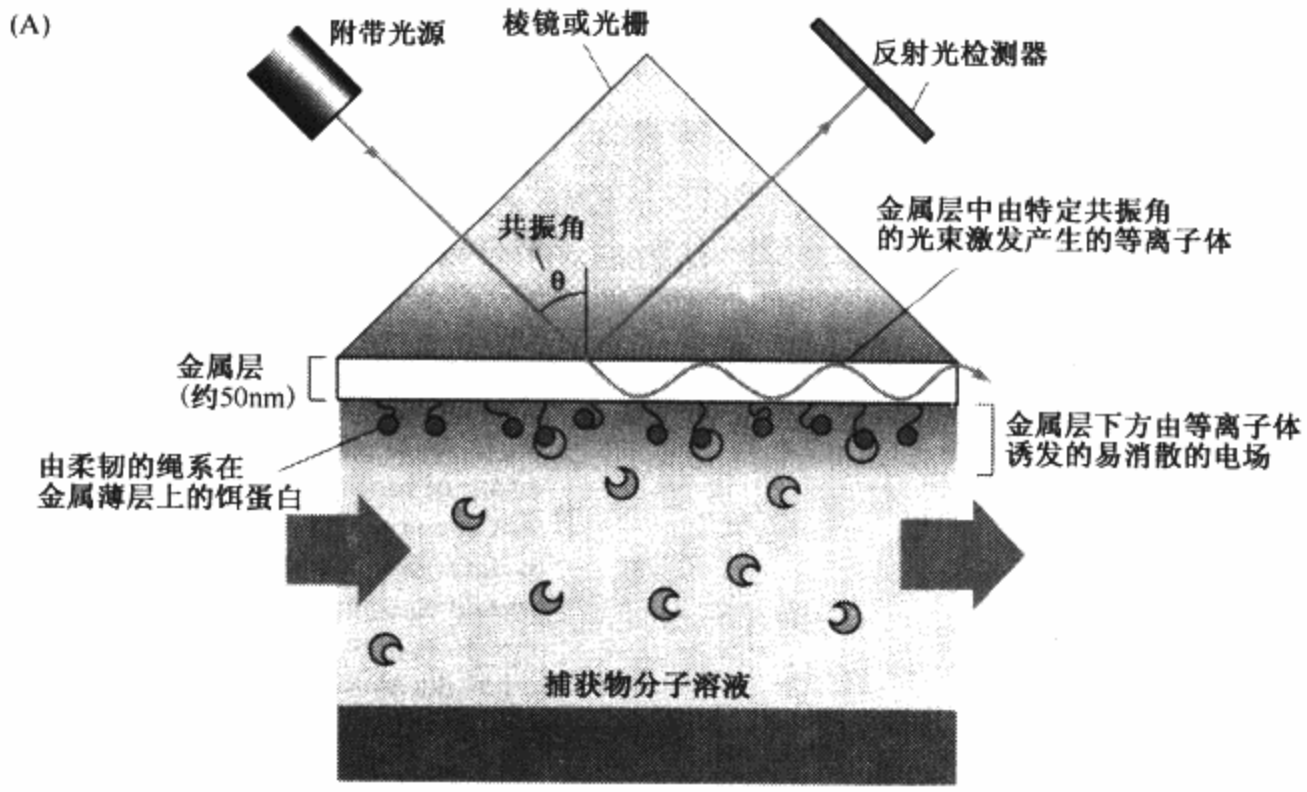


图 8-53　表面等离子体共振（SRR）

（A）SPR 可以通过监测可能的结合分子（绿色）水溶液和带有固定的饵蛋白（红色）的生物传感器间的界面上光束的反射来检测结合作用。（B）捕获蛋白溶液流过固定的饵蛋白。捕获物和饵蛋白的结合产生了一个可测量的共振角变化。这些实时监测的改变，反映了分子复合物的结合和分离

DNA 足迹法可以显示蛋白质在 DNA 分子上结合的位置

目前我们已经集中讨论了蛋白质-蛋白质相互作用。但是，一些蛋白质通过结合到 DNA 上行使功能。这些蛋白质大多数通常结合到基因编码区域之外的一段 DNA 调节序列，对决定在特定的细胞中哪些基因具有活性有核心作用。

在分析这样的蛋白质如何行使功能的过程中，鉴别它所结合的特异核苷酸序列是十分重要的。为了实现这一目标所采用的方法叫做 **DNA 足迹法**。首先，纯化分离出将一条一端标记了^{32}P 的 DNA 片段［图 8-24(B)］；然后用核酸酶或者其他化学物质切割这一分子，在 DNA 上产生随机的单链切口。变性分离 DNA 双链之后，得到的片段从标记端开始在凝胶中进行分离，并进行放射自显影检测。比较 DNA 结合蛋白质存在时进行 DNA 切割得到的条带图与

没有结合蛋白质时进行 DNA 切割得到的条带图。当蛋白质存在时，它会覆盖它的结合位点处的核苷酸，保护它们的磷酸二酯键，防止切割。结果就无法得到终止于结合位点的标记片段，在凝胶上形成空缺，叫做“足迹”（图 8-54）。相似的方法可用于测定蛋白质的 RNA 结合位点。

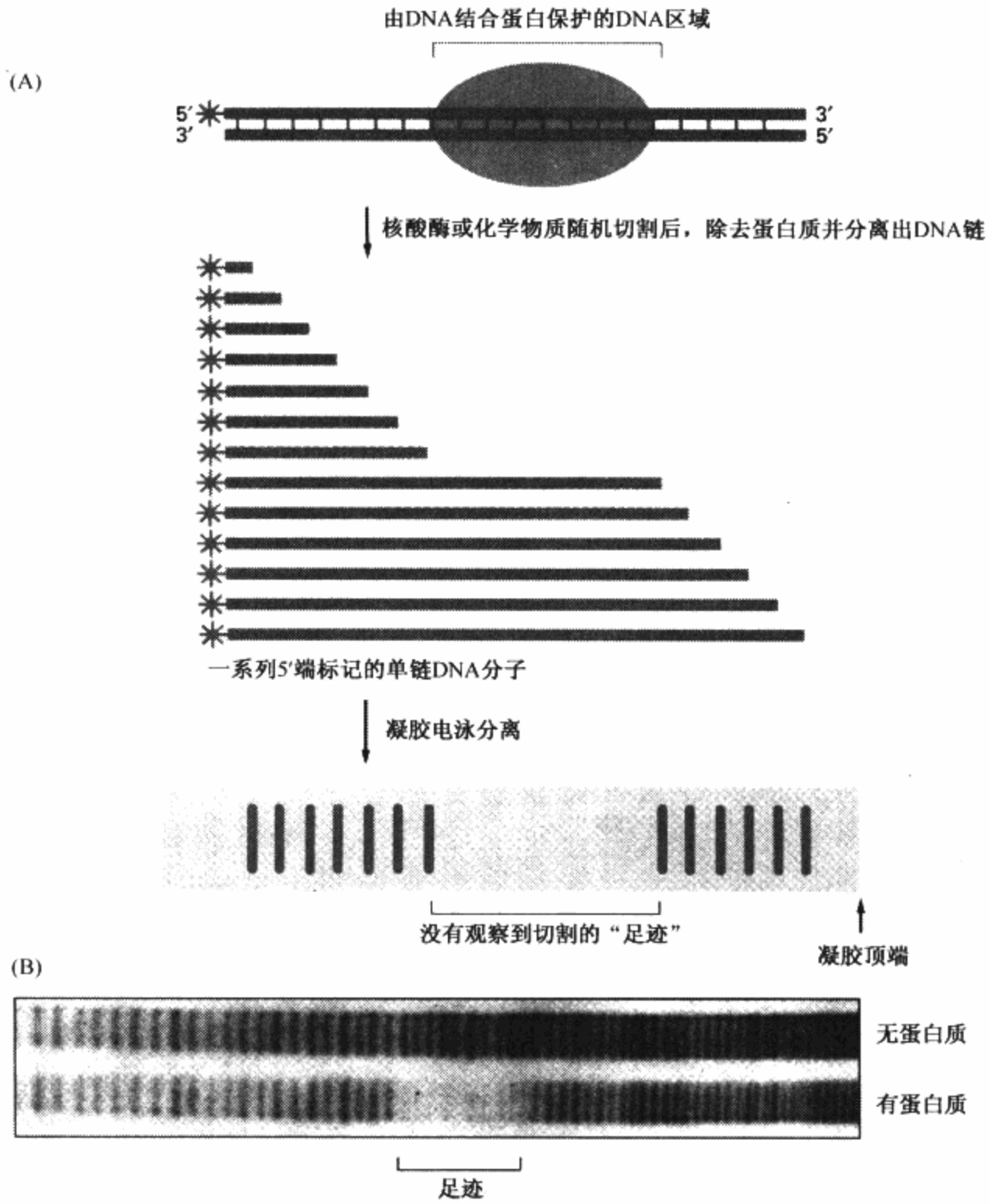


图 8-54　DNA 足迹技术

（A）这一技术需要一个一端被标记的 DNA 分子［图 8-24(B)］。图示的蛋白质与一段长度为 7 个核苷酸的特异 DNA 序列紧密结合，因此保护了这 7 个核苷酸，防止它们被切割物切割。如果在没有 DNA 结合蛋白存在时进行同样的反应，就会在凝胶上看到一个完整的条带阶梯（未在图上表示）。（B）一个用于检测一种人类蛋白质的结合位点的足迹法实例，这种蛋白质激发特异真核基因的转录。这些结果将结合位点定位于 RNA 合成起点的上游 60 个核苷酸。切割物是一个小的含铁的有机分子，可以几乎相同的频率切割每一个磷酸二酯键［（B）由 Michele Sawadogo 和 Robert Roeder 提供］

小结

许多有效的技术应用于研究蛋白质的结构和功能。为了测定在原子分辨率水平上的蛋白质三维结构，大的蛋白质需要进行结晶并通过 X 射线衍射技术进行研究。溶液中的小蛋白质的结构可以用核磁共振分析来测定。由于有相似结构的蛋白质常常具有相似的功能，有时可以通过搜索具有相似氨基酸序列的已知蛋白质来预测蛋白质的生化活性。

研究蛋白质功能的其他线索来自对它的亚细胞分布的检测。将蛋白质与绿色荧光蛋白（GFP）等分子标记进行，可以跟踪蛋白质在细胞内的运动。进入细胞核并且与 DNA 结合的蛋白质可以进一步用 DNA 足迹法进行分析，这种技术可用来测定蛋白质调节基因转录时结合到哪些调节序列。

所有蛋白质都要通过与其他蛋白或者分子结合来行使功能，有许多方法可用于研究蛋白质-蛋白质相互作用和鉴别可能的蛋白质伙伴。利用靶蛋白的抗体进行蛋白质亲和层析或者免疫共沉淀，可以对相互作用的蛋白质进行物理分离。其他技术，如双杂交系统或噬菌体展示，可以同时分离相互作用蛋白质和编码它们的基因。然后就可以测定蛋白质或者其相应基因的序列，从而确定从这些方法中得到的蛋白质的身份。

研究基因的表达和功能

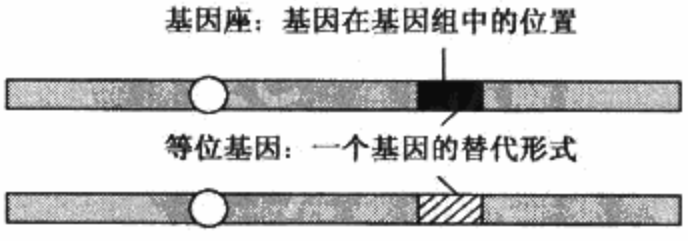
人们最终想要确定的是基因和它编码的蛋白质如何在完好的生物体中行使功能。尽管听上去是违反直觉的，找出一个基因做什么的最直接的方法是看当失去这一基因时生物体会怎么样。研究具有核苷酸序列改变或者缺失的突变生物体是生物学中的一项费时的实践。由于突变会中断细胞中的系列进程，突变体常常掌握着理解基因功能的钥匙。在**遗传学**重要领域的经典途径中，人们就是从分离突变体开始的，这些突变体具有有趣的或者异常的外观，如具有白眼或卷翅的果蝇。从**表型**（个体的外观或行为）追溯，就可以确定生物体的**基因型**，相应于这一特征的基因形式（专题 8-1）。

专题 8-1　经典遗传学回顾

基因和表型

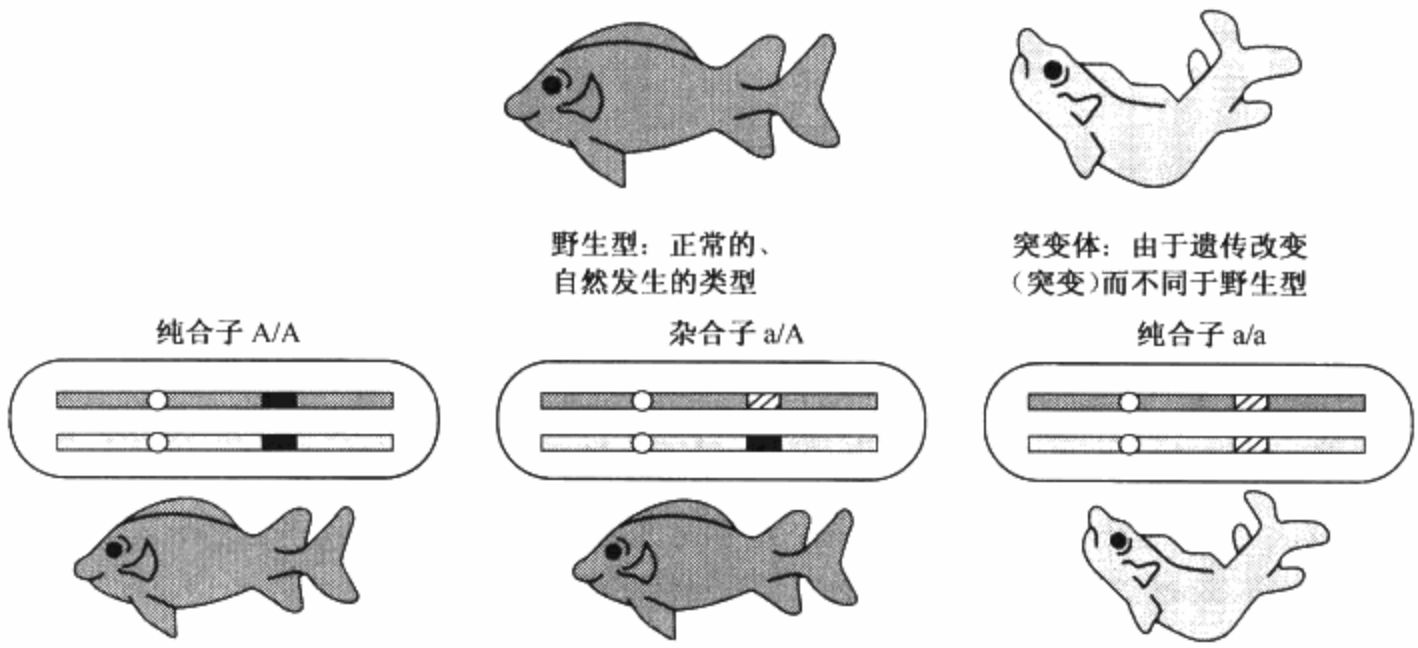
基因：可遗传的功能单元，通常对应于编码一个单独蛋白质的 DNA 片段。

基因组：一个生物体的一套基因。



基因型：形成个体基因组的特定等位基因构成。

表型：个体可见的特征。

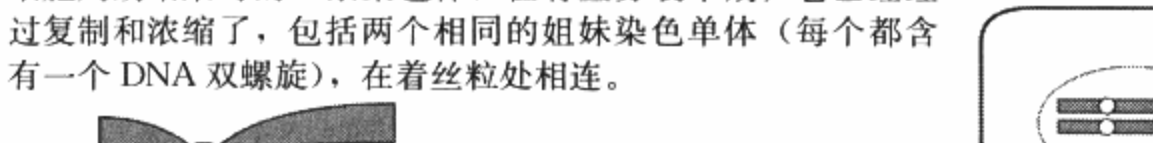


等位基因 A 是显性的（相对于 a）；等位基因 a 是隐性的（相对于 A）

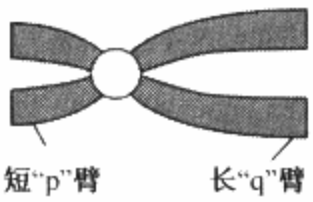
在上面的例子中，杂合子的表型与一种纯合子的表型相同；如果杂合子表型与两种纯合子表型都不相同，这两个等位基因就被称为是共显性的。

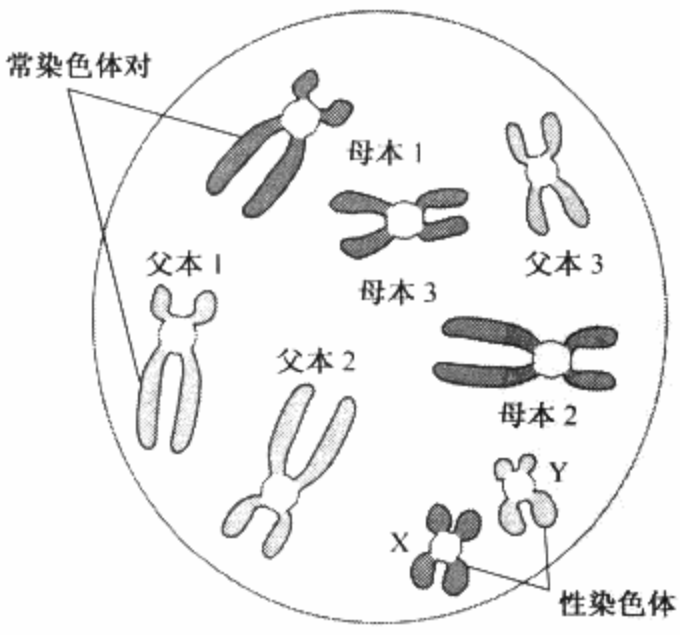
染色体

细胞周期开始时的一条染色体，在 G_1 期；单一的长条代表一条长的 DNA 双螺旋。



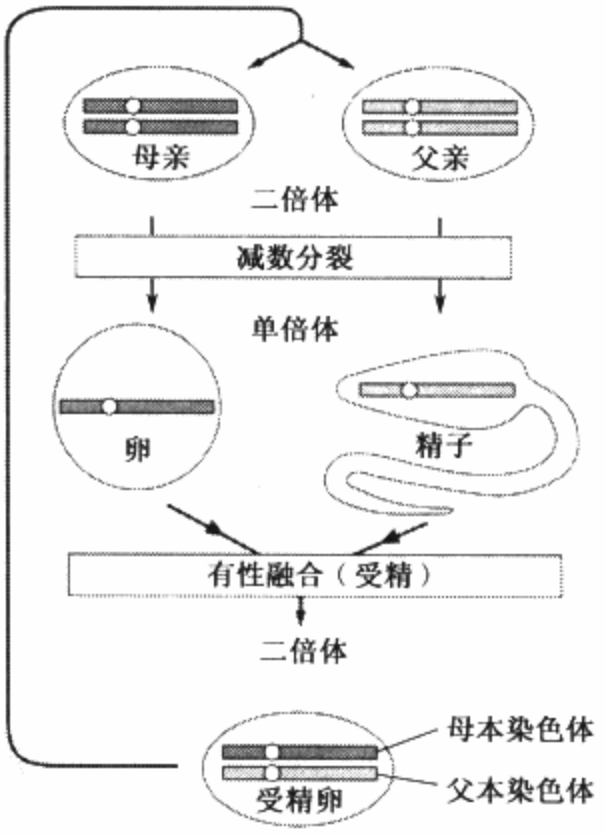
细胞周期结束时的一条染色体，在有丝分裂中期；它已经经过复制和浓缩了，包括两个相同的姐妹染色单体（每个都含有一个 DNA 双螺旋），在着丝粒处相连。





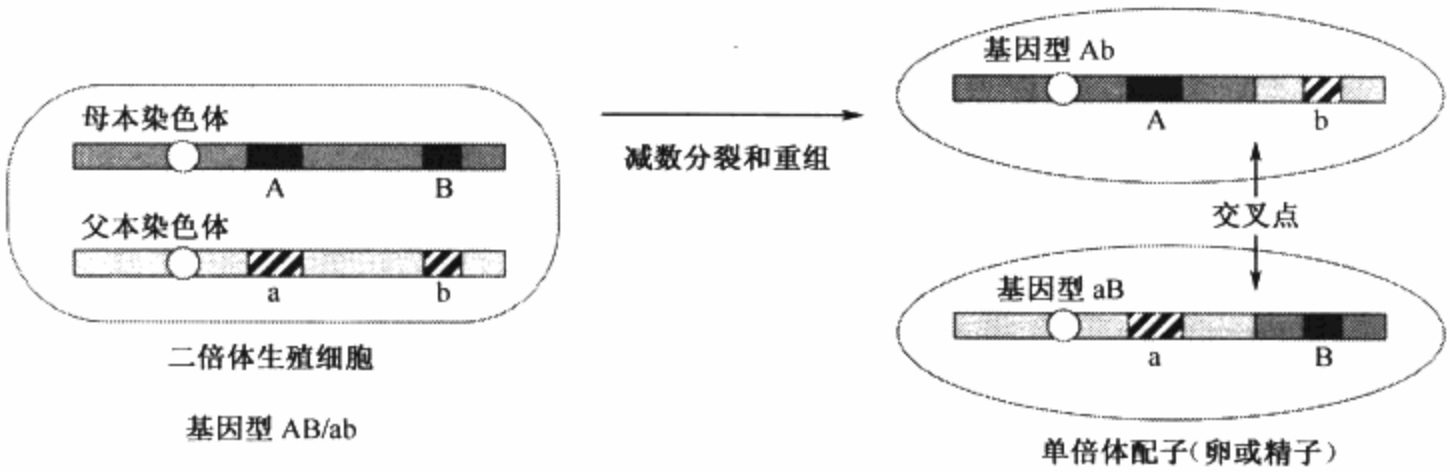
一套正常的二倍体染色体的中期涂片，通过破裂中期细胞并将散开的染色体染色进行制备。在图中所示的例子中，由 3 对常染色体（从两个亲本对称遗传来的染色体，不考虑性别）和两条性染色体（一条 X 染色体来自母本，一条 Y 染色体来自父本）。性染色体的数目、类型和它们在性别决定中的作用，以及常染色体对的数目在不同种类的生物体中各不相同。

有性生殖的单倍体-二倍体循环



为了简单起见，只用了一对染色体来表示这一周期。

减数分裂和遗传重组



二倍体生殖细胞

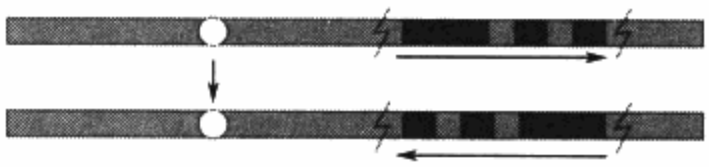
基因型 AB/ab

单倍体配子(卵或精子)

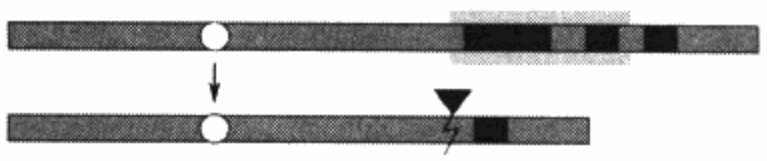
两个基因座在同一条染色体上相距的越远，通过它们之间一点的交叉将它们分开的概率就越大。如果两个基因在配子中重新分配的概率为 $x\%$，那么它们在遗传图谱的一条染色体上相距 x（xcM)。

突变类型

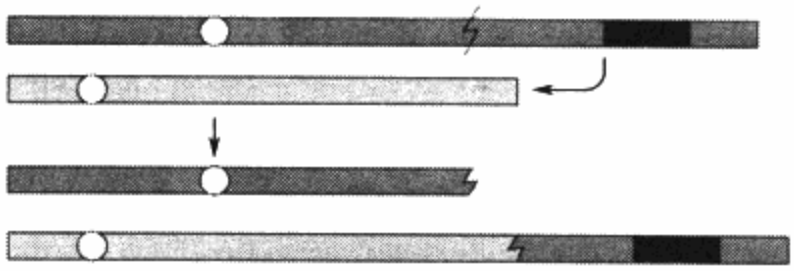
点突变：基因组中的单一位点图示，对应于一个单独的核苷酸对或一个单独基因的很小一部分



倒置：颠倒染色体的一个片段



缺失：删除染色体的一个片段



易位：一条染色体的一个片段脱落下来，连接到另一条染色体上

致死突变：使得发育中的生物体未成熟即死亡。

条件突变：只在特定条件下产生表型效应，这一特定条件叫做限制性条件。在其他条件（允许条件）下，看不到表型效应。对于一个温度敏感突变而言，限制性条件一般为高温，允许条件一般为低温。

功能缺失突变：降低或彻底破坏了基因的活性。这是最普遍的一类突变。功能缺失突变通常是隐性的，生物体通常可以正常行使功能，只要它确保有至少一个拷贝有用的基因。

无效突变：一种使得基因功能完全丧失的功能缺失突变。

功能获得突变：增加了基因的活性，或者使它在不适合的环境下有活性；这些突变通常是显性的。

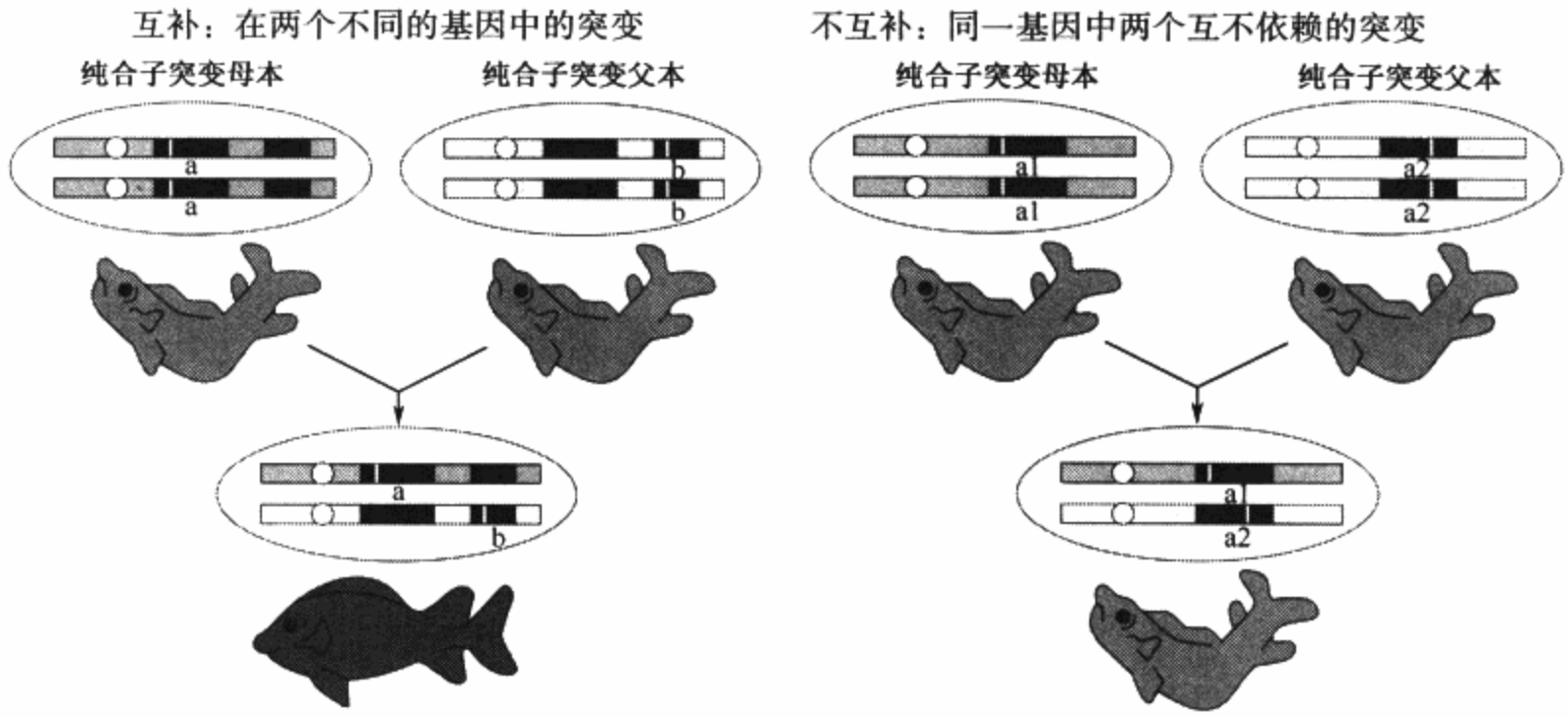
显性失活突变：阻断基因活性的显性突变，即使有一个正常拷贝的基因存在时，也产生功能缺失表型。当突变基因产物干扰了正常基因产物的功能时，这种现象就发生了。

抑制突变：抑制另一突变的表型效应，从而使双突变体看上去是正常的。基因内抑制突变位于第一个突变作用的基因内；基因外抑制突变位于第二个基因——这一基因的产物经常可以和第一个基因产物直接相互作用。

两个基因还是一个?

给出两种产生同一表型的突变，我们怎样才能指出它们是否是同一基因上的突变呢？如果突变是隐性的（通常是这样），可以通过互补测验来找到答案。

在最简单的互补测验中，某一突变的纯合子个体与另一突变的纯合子个体交配，子代的表型给出了问题的答案。



今天，众多的基因组计划每天向公共的数据库中添加成千上万的核苷酸序列，对基因功能的探索常开始于一段 DNA 序列。这里面临的挑战是将序列转译为功能。在本章前面讨论过的一种方法是，搜索与新基因编码蛋白质有相似氨基酸序列的已知蛋白质，并利用前面部分已经讲过的方法来进一步探索基因的功能。但是要直接解决一个基因如何在细胞或生物体中行使功能的问题，最有效的方法是研究缺失这一基因或者表达其变化形式的突变体。确定在这种突变体中哪种细胞进程被中断或者受到损害，常常可以帮助人们了解一个基因的生物学功能。

在这一部分，我们会描述许多不同的用于测定基因功能的方法，无论是从一段 DNA 序列开始，还是从一个具有有趣表型的生物体开始。我们首先从经典的遗传性方法入手来研究基因和基因的功能。这些研究开始于遗传筛选，来分离感兴趣的突变体，然后鉴别对应于观察到表型的基因或基因群。然后我们回顾一下反向遗传学技术，即从基因或基因序列入手来确定其功能。这种方法通常需要一些聪明的预测，搜索同源序列并确定基因在何时何处表达，同时产生突变生物体并鉴定它们的表型。

经典途径从随机突变开始

在基因克隆技术出现之前，大多数基因是通过基因突变中断生物进程来鉴定的。这种经典的遗传学方法（鉴定与突变表型相应的基因）最容易在迅速繁殖且易进行遗传操作的生物体中施行，如细菌、酵母、线虫和果蝇。虽然有时也可以在检查非常大的种群（数千或者数万个个体）时发现自发突变，但是用损害 DNA 的物质产生突变，可以使分离突变体变得更有效。用诱变剂处理生物体，可以快速产生大量的突变体，然后可以筛选出一个感兴趣的缺陷，不久我们会讨论。

化学或辐射诱变的一种替代过程叫做插入诱变。这种方法基于这样一个事实：如果插入片段中断了一个基因或者它的调控序列，随机插入基因组的外源 DNA 就可以产生突变。已知序列的插入 DNA 可以作为分子标签，帮助后续的被中断基因的鉴别和克隆（图 8-55）。在果蝇中，利用可转移的 P 因子来使基因失活，带来了果蝇基因功能研究的革命。转座因子（表 5-3）也用于在细菌、酵母和开花植物拟南芥中产生突变。可以将自身基因拷贝入宿主基

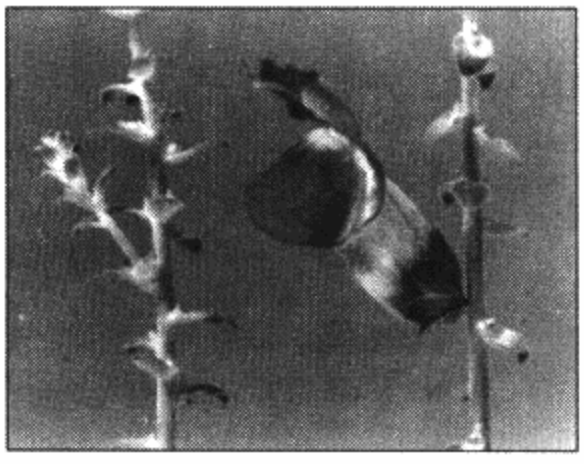


图 8-55　金鱼草的插入突变体

因组的反转录病毒（图 5-73），已经用于中断斑马鱼和老鼠中的基因。

这些研究很适合于仔细研究线虫和果蝇中的生物进程，但是我们如何研究人类的基因功能呢？不同于我们已经讨论过的生物体，人类不能迅速繁殖，而且不能故意用诱变剂处理。此外，如 DNA 复制等重要进程具有严重缺陷的任何人类，都在出生之前就死亡了。

对于如何研究人类基因，有两种答案。第一种，由于基因和基因功能在进化过程中都是高度保守的，研究不太复杂的模式生物可以揭示人类相似的基因和进程。然后，相应的人类基因就可以在培养的人类细胞中进一步研究。第二种，许多突变是非致死性的，例如，组织特异的溶菌酶或细胞表面受体缺陷，这些突变已经在人类种群中自发出现。分析受到影响的个体的表型，结合对他们的培养细胞的研究，已经提供了对人类重要细胞功能的独特知识。尽管这种突变很罕见，但仍然可以有效地发现它们，因为人类有一个特性：突变个体通过寻求特殊的医学治疗会使自己被注意到。

遗传学筛选鉴别突变体在细胞进程中的缺陷

一旦在模式生物如酵母或果蝇中产生了一系列突变，通常必须检查数千种个体，来寻找感兴趣的变化表型。这种搜索叫做**遗传学筛选**。由于得到感兴趣基因中的突变依赖于在随机诱变中这种基因被失活或者以其他方式突变的可能性，基因组越大，某一特殊基因被突变的可能性越小。因此，生物体越复杂，就需要检查越多的突变体，以避免漏掉某些基因。待筛选的表型可以简单也可以复杂。简单的表型易被检测，如一个代谢缺陷，生物体将不能在某一种特殊氨基酸或营养物缺乏的条件下生长。

更加复杂的表型，如引起学习或者记忆方面缺陷的突变，可能需要更复杂精细的筛选（图 8-56）。但是，即使用于研究复杂生理学系统的遗传学筛选，在设计时也应当尽可能的简

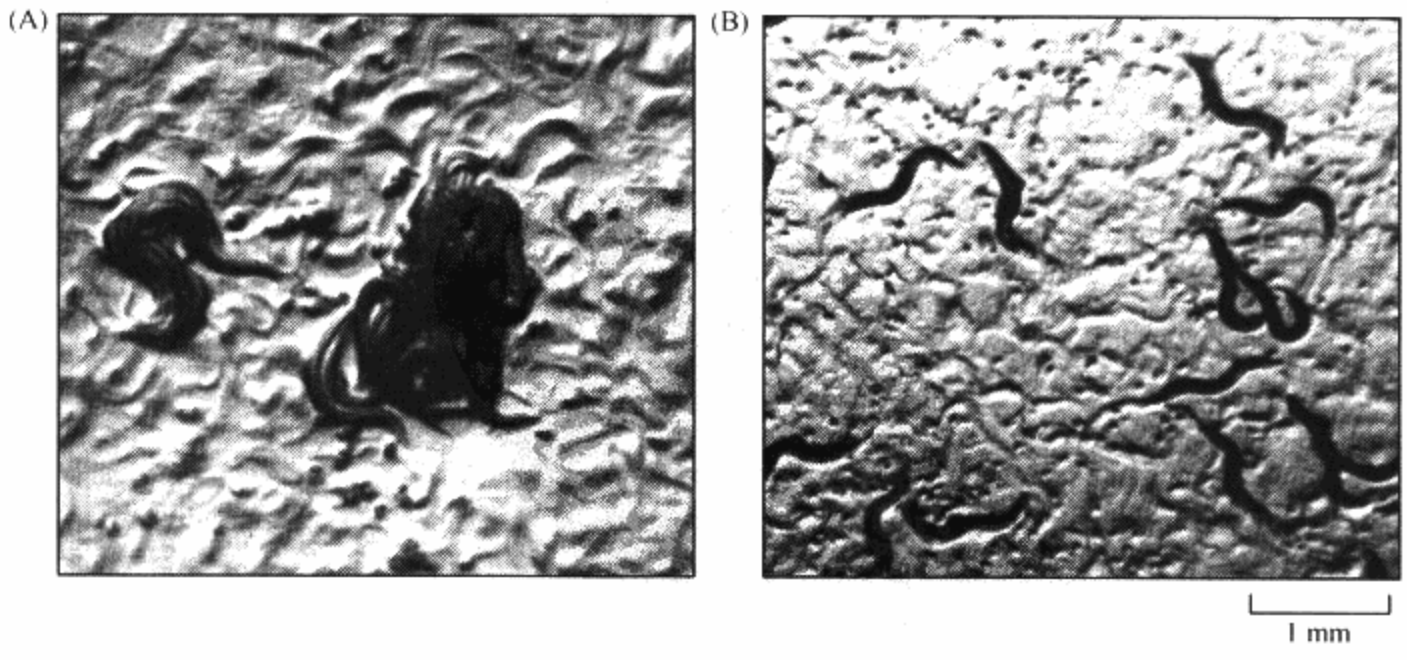


图 8-56　筛选可以检测影响动物行为的突变

（A）野生型线虫进行群体进食。线虫四处游动，直到找到它们的邻居并开始进食。（B）突变的动物自己进食

单，如果可能，应当允许同时进行大量突变体的检查。举个例子，一个特别好的筛选设计来寻找在斑马鱼视觉形成过程的基因。这一筛选是监测鱼对运动的反应，其根据是行为上的改变。野生型鱼往往按照感觉到的方向游动，而具有视觉系统缺陷的突变体向着随机的方向游动，这是一种很容易检测的行为。在这一筛选中发现的一种突变命名为 *lakritz*，在这种突变中，有 80%的将视觉信号从眼睛传播到脑的视网膜神经节缺失。由于斑马鱼视网膜的细胞组织状态可以代表所有脊椎动物的，对这些突变体的研究也可以帮助我们理解人类的视觉发生过程。

由于在基本的细胞过程，例如，RNA 合成加工或细胞周期调控中需要基因的缺陷通常是致死的，这些基因的功能通常在温度敏感性突变体中进行研究。在这些突变中，突变基因的蛋白质产物在中等的温度可以正常发挥功能，但是温度的微小上升或降低就会使其失活。因此，可以通过改变温度来在实验室中简单地打开或关闭这一异常性。含有对非允许温度下生存至关紧要的基因的温度敏感突变的细胞，仍然能在正常温度或者允许温度下生长（图 8-57）。在这种突变中的温度敏感基因通常含有一个引起蛋白质产物的精细变化的点突变。

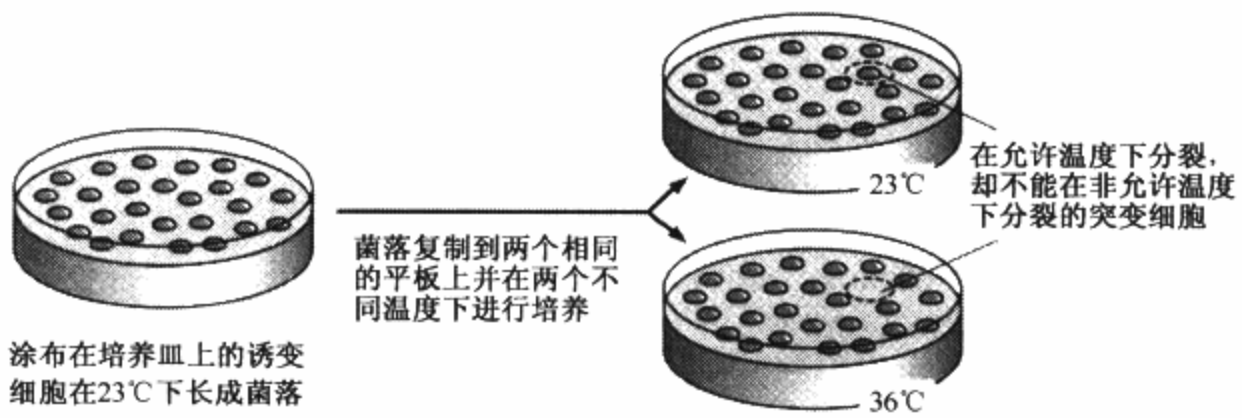


图 8-57　筛选温度敏感的细菌或酵母突变体

诱变的细胞在允许温度下涂布，得到的菌落通过复制涂布转移到两个相同的培养皿上：一个在允许温度下培养，另一个在非允许温度下培养。含有在对增殖至关紧要的基因的温度敏感突变的细胞可以在正常的、允许温度下分裂，却不能在提高的、非允许温度下分裂

许多温度敏感突变体是在编码 DNA 复制所需细菌蛋白质的基因中分离得到的，通过筛选诱变的细菌种群来寻找从 30℃升温到 42℃后停止形成 DNA 的细胞。这些突变体接着用于区别和鉴定相应的 DNA 复制蛋白（讨论见第 5 章）。温度敏感突变体也导致了许多细胞周期调节相关蛋白和酵母中分泌途径的移动蛋白的鉴定（专题 13-1）。利用相似的筛选方法研究了细菌和酵母的基础代谢途径相关酶的功能（讨论见第 2 章），并且发现了许多与果蝇胚胎有序发育有关的基因产物（讨论见第 21 章）。

互补测验可以揭示两个突变是否发生在同一个基因内部

大范围的遗传学筛选可以发现许多表现出同一表型的不同突变。这些突变可能发生在作用于同一进程的不同基因上，或者代表同一基因的不同突变。那么，我们怎样判断产生同一表型的两种突变是发生在同一基因内呢，还是发生在不同基因上？如果突变是隐性的，举个例子，如果它们代表一个特别基因的失去功能突变，那么可以用互补测验来确定突变是在同一基因还是在不同基因。互补测验的最简单方式是，一个个体是某一突变的纯合子，即它有待测突变两个相同的等位基因，将它与另一突变的纯合子个体交配。如果两个突变在同一基因内，子代将表现突变表型，因为它们仍然没有待测基因的正常拷贝（专题 8-1）。相反地，如果突变落在不同的基因上，产生的后代就会表现正常的表型。它们得到了每个基因的一个

正常拷贝（和一个突变拷贝）。于是突变就互相补充，恢复了正常表型。在遗传学筛选过程中进行的突变体的互补测验表明，例如，酵母消化半乳糖需要 5 个基因，大肠杆菌建立一根有功能的鞭毛需要 20 个基因，48 个基因与 T4 噬菌体的病毒颗粒组装过程相关，上百个基因与从受精卵到成熟线虫的发育过程相关。

一旦鉴定出与一个特殊生物进程相关的一系列基因，接下来就要测定这些基因发挥功能的顺序。测定基因何时活动可以促进再现整个遗传学或生物化学途径，这种研究也已经成为我们理解代谢、信号转导、其他发育和生理进程的中心。其实，解开基因行使功能的顺序需要仔细地鉴定每个不同基因的突变所产生的表型。举个例子，假如一堆基因的突变都引起早期胚胎发育的细胞分裂停止，对每个突变的精密检查就可能揭示某些基因作用得十分早，阻止了受精卵分裂成两个细胞；而其他基因可能允许早期的细胞分裂却组织胚胎进入囊胚期。

为了测验对基因功能顺序的预测正确与否，可以让生物体具有两个不同基因的突变。如果这些突变影响同一进程的不同步骤，那么这种双突变体应该具有和在进程中最早作用的突变的表型一致。举个例子，酵母的蛋白质分泌过程就是通过这一方法得以解释的。这一过程的不同突变产生了在内质网或高尔基体中异常的蛋白质聚集。当一个细胞被改造成具有阻断蛋白质在内质网中加工和阻断蛋白质在高尔基体中加工的两个突变时，蛋白质沉积在内质网中。这就表明蛋白质在运送到高尔基体之后，进行分泌之前，必须经过内质网（图 8-58）。

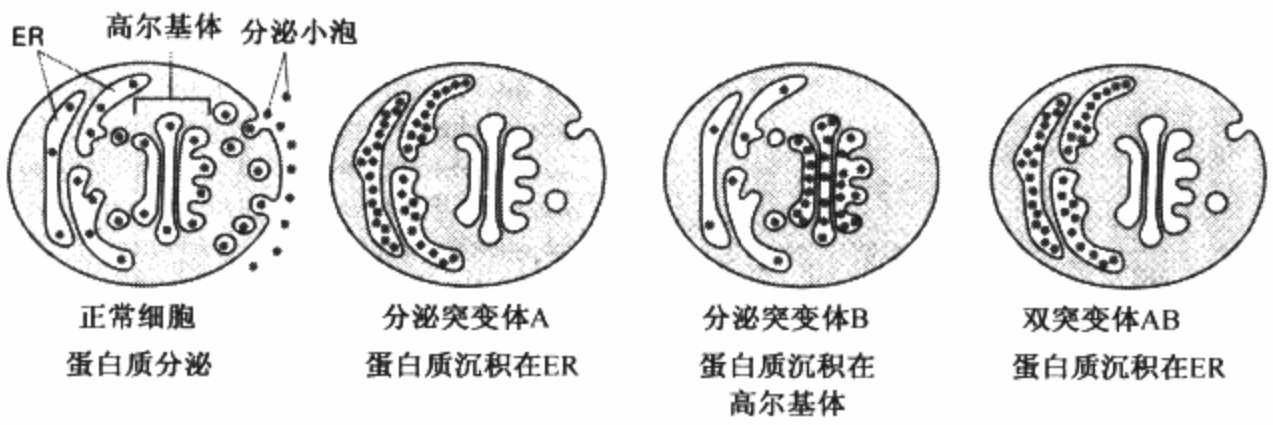


图 8-58　用遗传学测定基因功能的顺序

在正常细胞中，蛋白质被装入小泡，小泡与细胞质膜融合，将其内含物释放到细胞外基质。在分泌突变体 A 中，蛋白质沉积在内质网。在不同的分泌突变体 B 中，蛋白质沉积在高尔基体内。在双突变体 AB 中，蛋白质沉积在内质网；这表明突变体 A 缺陷的基因比突变体 B 缺陷的基因在分泌过程中发生作用要早

基因定位可以通过连锁分析

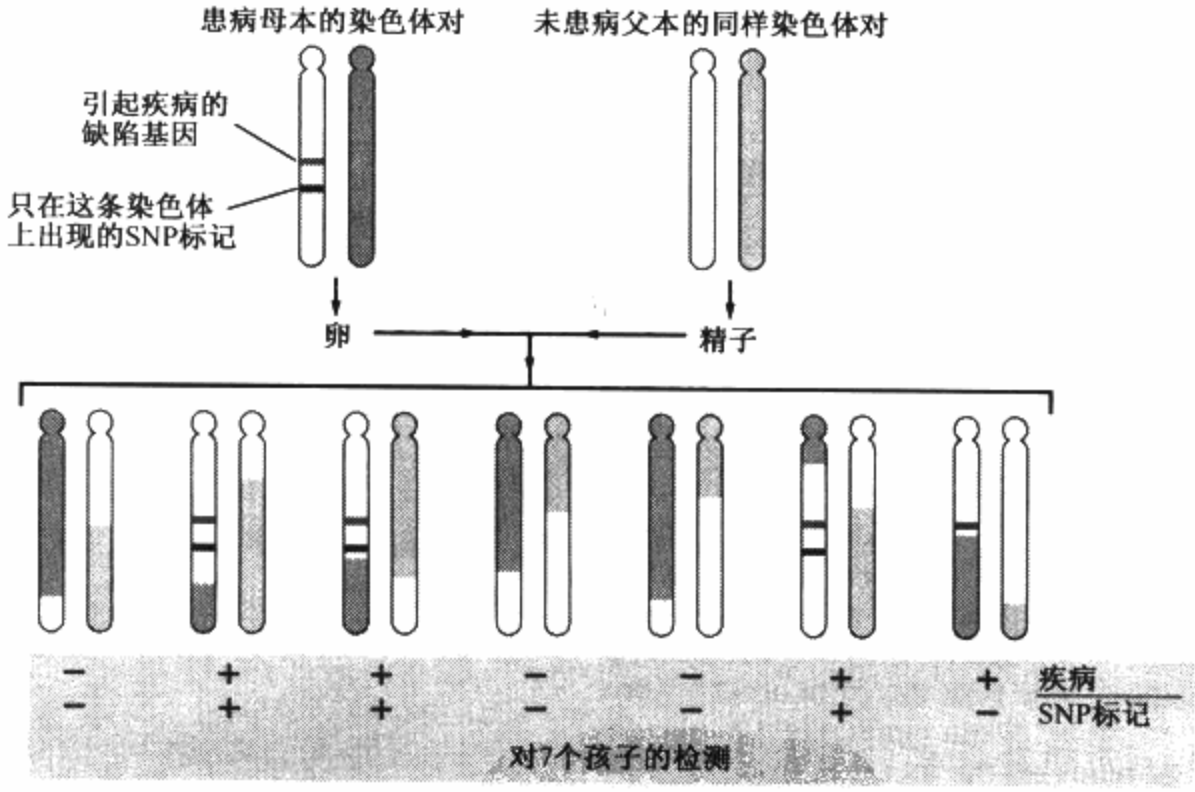
得到突变之后，下一步就是鉴定与改变表型相关的基因或基因群。如果最初使用的是插入诱变的方法，定位被打断的基因就十分简单了，只需要搜集和扩增含有插入因子（如转座子或反转录病毒）的 DNA 片段，并测定侧翼 DNA 的核苷酸序列。然后用这一序列搜索 DNA 数据库，来鉴定被插入因子打断的基因。

如果使用损害 DNA 的化学物质来产生突变体，鉴定失活的基因常常更费力一些，可以通过不同的方法完成这一过程。一种方法是，首先测定基因在基因组中的位置。为了定位一种新发现的基因，首先要通过评估这个基因与基因组中其他已知基因的距离，来测定它在染色体上的大约位置。通常使用连锁分析的方法来估计基因位点之间的距离，这一技术依据的事实是在同一染色体上互相靠近的基因趋向于一起遗传。基因之间离得越近，它们一起传给

子代的可能性就越大。但是，即使是紧密连在一起的基因，在减数分裂过程中仍有可能通过重组而被分开。两个基因位点间距离越大，它们通过交叉被分开的机会就越大（专题 8-1）。通过计算两个基因间的重组频率，可以测定它们之间的大概距离。

由于基因并不总是相距足够近得可以精确测定它们的位置，因此连锁分析常常依赖于基因组中的物理标记，来估计一个未知基因的位置。这些标记一般是核苷酸片段，具有已知的序列和基因组定位，并且至少有两种等位基因形式。例如，单核苷酸多态现象（SNP）是种群的个体中一个或多个核苷酸不同的短序列，可以通过杂交技术进行检测。许多这样的物理标记散布在染色体上，许多生物体中都搜集到了这样的序列，人类中有多于 10^6 个。如果这些标记分布得足够密集，人们就可以通过连锁分析检测与突变表型紧密共遗传的一个或多个 SNP，来缩小一个基因的可能定位的染色体区域，使其定位到一段只含有几个基因序列的区域上。这些基因就是候选基因，可以直接测定它们的结构和功能，来确定哪个基因是和最初的突变表型相关的。

连锁分析同样可用于鉴定与遗传性的人类疾病相关的基因。这种研究需要收集受到这种疾病影响的大量家族的 DNA 样品。然后，检查这些样品中似乎与疾病基因紧密相连的物理标记（如 SNP），这些疾病基因总是通过患者的个体遗传，而不通过他们不受影响的亲戚。然后就像上面所述的那样定位这种疾病基因（图 8-59）。例如，囊肿性纤维化和亨廷顿舞蹈症的基因都是通过这种方法发现的。



结论：致病基因在 75%的子代中与 SNP 标记共遗传自母本。如果在其他被检查的家族中发现同样的关系，致病基因就可以定位到这一染色体上这种 SNP 的附近。注意一个在同一染色体上距离基因很远的或与感兴趣的基因定位在不同染色体上的 SNP，共遗传的概率只有 50%

图 8-59　遗传学连锁分析利用 DNA 上的物理标记来寻找人类基因

在这个例子中，利用一种 SNP 标记研究人类的一种特异表型的共遗传（这里是一个遗传疾病）。如果遗传这一疾病的个体几乎总是遗传一种特异的 SNP 标记，那么引起这种疾病的基因和 SNP 就很可能在同一染色体紧密接近，如图中所示。为了证明一个观察到的连锁具有统计意义，需要检查数百个个体。注意连锁不会是绝对的，除非 SNP 标记位于这一基因自身。因此，偶然地，在卵或精子形成是通过减数分裂的交叉过程中，SNP 会与疾病基因分开：这一情况发生在图中的最右边的例子。当处理一个已经测序的基因组时，用原来的 SNP 两边的 SNP 不断重复这一流程，直到找到 100%的共遗传性

寻找同源性可以帮助预测基因的功能

一旦基因被鉴定了，常常可以通过确定其已知功能的同源基因来预测它的功能。像我们之前讨论过的，数据库中含有来自各种各样生物体中的核苷酸序列，包括许多微生物、线虫、拟南芥、果蝇和人类的完整基因组序列，可以搜索数据库来寻找和未鉴定的目标基因相似的序列。

当分析一个新测序的基因组时，这一搜索可以作为首选的尝试，来指派尽可能多的基因的功能，这一过程叫做评注。然后进行进一步的遗传学和生物化学研究，来确定基因是否编码具有预测功能的产物，之后我们会讨论。同源性分析并不总是能揭示关于功能的信息：在酵母基因组中，有 30%的未鉴定的基因可以通过同源性分析来指派可能的功能；10%具有功能未知的同源物；另外的 30%在任何现有的数据库中都找不到同源物（剩下 30%的基因在酵母基因组测序之前就已经被鉴定了）。

在某些情况下，同源性搜索找到的生物体 A 中的基因，它产生的蛋白质在不同生物体中是与第二种蛋白质融合的，但是第二种蛋白质在生物体 A 中是由一个独立的基因产生的。例如，在酵母中，两个分离的基因编码两种与色氨酸合成相关的蛋白质；但在大肠杆菌中，这两种基因融合成一个基因（图 8-60）。酵母中的这两种蛋白质对应于一个细菌蛋白质的两个结构域，意味着它们可能具有功能上的相关性，而且可能在同一个蛋白质复合物中发挥作用。更普遍地，这种方法用于建立基因间的功能联系，而在大多数生物体中这些基因在基因组中离得很远。

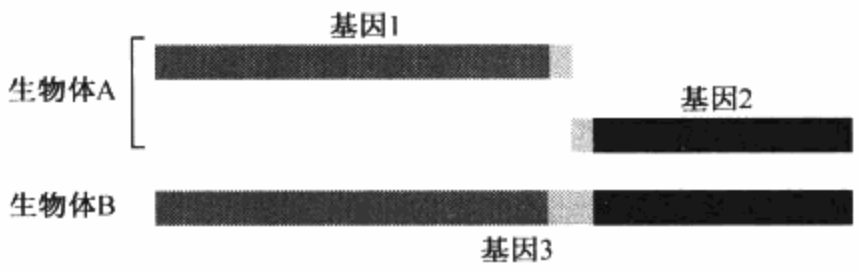


图 8-60　结构域融合揭示功能相关的基因间的关系

在这个例子中，生物体 A 的基因 1 和基因 2 的功能性相互作用，是通过推断生物体 B 中单一基因（基因 3）的同源性结构域的融合得到的

报道基因可以揭示基因何时何处表达

通常可以通过检查基因在细胞或整个生物体中何时何处表达来获得基因功能的线索。可以利用报道基因代替所研究基因的一部分来测定基因表达的模式和时间。大多数情况下，报道基因的表达可以通过追踪其荧光或蛋白质产物的酶活性来监测。

如第 7 章所详细讨论的，基因表达由位于编码区上游或下游的调节 DNA 序列所控制，它们一般不被转录。这些调节序列控制哪个细胞在什么条件下表达基因，也可以促进报道基因的表达。用报道基因简单地替代目标基因的编码区，并将这种重组的 DNA 分子导入细胞。报道基因产物的水平、时间和细胞特异性反映了属于原有基因的调节序列的作用（图 8-61）。

前面讨论的其他几种技术，也可以用于测定基因的表达模式。杂交技术，如 Northern 分析（图 8-27）和检测 RNA 的原位杂交（图 8-29）可以揭示基因什么时候在哪个组织被转录，产生了多少 mRNA。

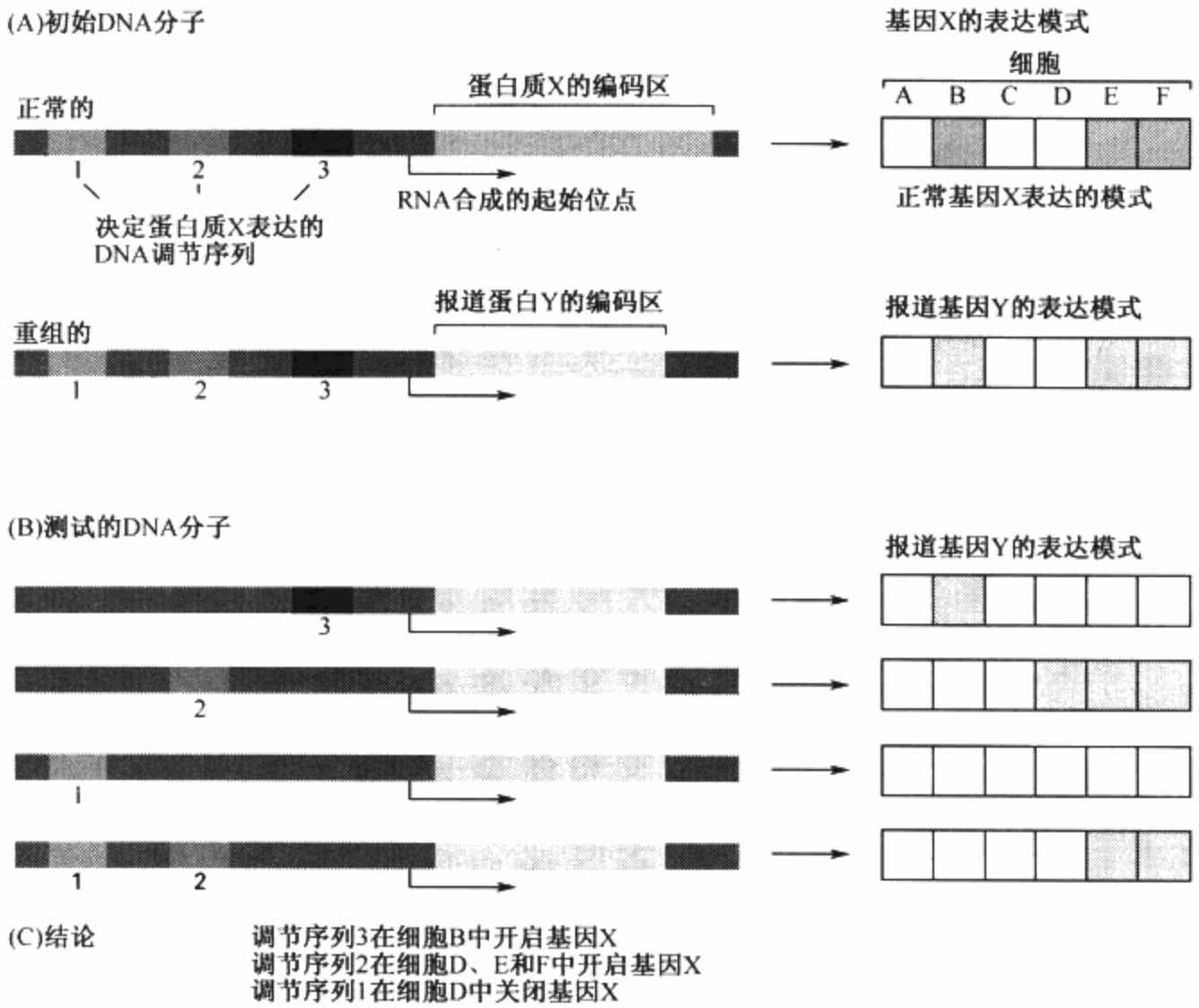


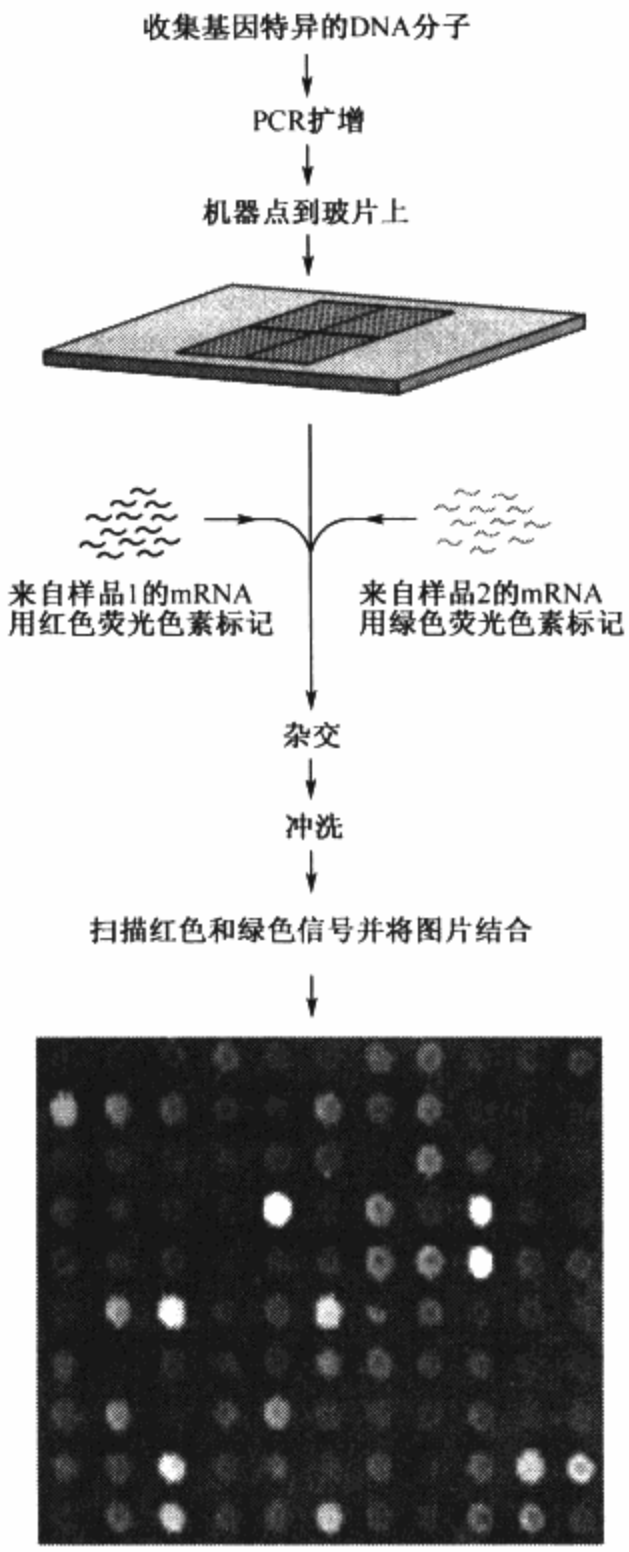
图 8-61　利用报道蛋白来测定基因表达的模式

（A）在这个例子中，蛋白质 X 的编码区被蛋白质 Y 的编码区替代。（B）多种含有候选调节序列的 DNA 片段加入到重组物中。将重组 DNA 分子转到不同类型的哺乳动物细胞中，然后测试它们的表达，表达的结果总结在（C）中。在真核细胞实验中，两种常用的报道蛋白是 β-半乳糖苷酶（β-gal）和绿色荧光蛋白（GFP）（图 9-44）。由于这些是细菌的酶，它们的存在可以通过简单灵敏的酶活测定进行监测，不会干扰宿主细胞的酶。图 7-39 显示了利用 β-gal 报道基因进行果蝇胚胎中眼睛基因调节序列活性监测的一个例子

芯片可以一次监测数千种基因的表达

目前我们已经讨论了可以一次监测一种基因表达的技术。这些方法许多是极费劳力的：构建报道基因或 GFP 融合需要进行 DNA 操作并将得到的重组分子转入细胞。即使 Northern 分析也被限制在琼脂糖凝胶上的样品数量的范围内。1990 年发展起来的 **DNA 芯片技术**带来了基因表达分析方法的革命，现在可以一次监测数千个基因的 RNA 产物。通过同时检查这么多基因的表达，我们可以开始鉴别和研究细胞生理学下的基因表达模式：我们可以看到细胞生长、分裂或者对激素或毒素发生反应时哪些基因开启（或关闭）。

DNA 芯片只比显微镜玻片多了散布于其上的大量 DNA 片段，每个片段含有一个可作为特异基因探针的核苷酸序列。最密的芯片可能在比邮票小的区域内含有上万种片段，允许数千种杂交反应平行地进行（图 8-62）。一些芯片是由机器人将 PCR 产生的 DNA 大片段点到玻片上得到的。其他的含有短的寡核苷酸，这些寡核苷酸是用类似于制备电脑芯片上的蚀刻电路的技术在玻璃干胶片表面合成的。每种情况下，芯片上每个探针的确切序列和位置都是已知的。因此，任何与芯片上探针杂交的核苷酸片段都可以通过检测它的结合位置来鉴定是哪种特异基因的产物。



芯片代表110种酵母基因表达的区域

图 8-62　利用 DNA 芯片同时监测数千种基因的表达

为了制备芯片，对应于每个基因的 DNA 片段由机器人点到玻片上，也有商品化的制好的芯片。在这个例子中，mRNA 是从两个不同的细胞样品中得到的，来直接比较它们基因表达的相对水平。这些样品转变为cDNA并且进行标记，一种用红色荧光色素，另一种用绿色荧光色素。标记的样品混合在一起，并在芯片上进行杂交反应。温育后，冲洗芯片，然后进行荧光扫描。所显示的芯片部分代表了 110 种酵母基因的表达，红色点说明样品 1 中的基因比样品 2 中的相应基因具有更高的表达水平；绿色点说明基因表达的水平在样品 2 中比样品 1 中高。黄色点说明基因在两个样品中表达水平相同。暗点说明在芯片这一位置的基因在每种样品都低表达或不表达。更多细节见图 1-45（摘自 J. L. DeRisi et al., *Science* 278：680～686，1997）

利用 DNA 芯片来监测基因表达，首先从所研究的细胞中提取 mRNA 并转化成 cDNA（图 8-34）。然后，用荧光探针标记 cDNA。芯片与这种标记的 cDNA 样品温育，就发生了杂交反应（图 8-62）。接着冲洗芯片，除去未紧密结合的 cDNA，用自动化的扫描激光显微镜来鉴别芯片上标记 DNA 片段的结合位置。然后，将芯片位置与点在这一位置的 DNA 样品中的特异基因进行匹配。

实验样品中的荧光 DNA（如用红色荧光染料标记）要与标记了另一种不同颜色荧光染料（如绿色）的参照样品 cDNA 片段混合。这样，如果感兴趣的细胞中的一种特异基因表达的 mRNA 量相对于参照样品有所提高，得到的点就是红色。相反地，如果基因表达相对于参照样品有所降低，点就是绿色。利用这种内在的参照物，就可以将基因表达的轮廓精确的列出来。

目前，DNA 芯片已经用于检查所有事件，从使草莓成熟的基因表达变化到不同类型的人类癌症细胞的基因表达标记（图 7-3）。含有代表酵母所有的 6000 种基因探针的芯片已经用于监测酵母从发酵葡萄糖移到乙醇中生长的过程中，在对突然移到热环境或冷环境时，以及它

们在细胞周期的不同阶段的基因表达变化。初步已经表明，当酵母用光培养基中的葡萄糖时，它们的基因表达模式发生标志性的变化：接近 900 种基因更加活跃地进行转录，而其他 1200 种活性降低。这些基因中，有大约一半还不知道它们的功能，尽管研究表明它们在某种程度上和酵母从发酵到呼吸转变的过程中的代谢改变有关。

基因表达的全面研究也对基因功能预测提供了有用的信息。前面讨论的如何鉴定一个蛋白质的相互作用伙伴，可以带来蛋白质功能的线索。类似的原理对基因也适用：有关基因功能的信息可以从和它具有相同表达模式的基因中推断出来。利用一种叫做群分析的技术，可以鉴定协同调节的基因系列。在不同条件下一起开启和关闭的基因群可能在细胞中一起工作：它们可能编码同一多蛋白质机器的零件蛋白质，或者与一个复杂的协同作用（如 DNA 复制或 RNA 剪接）相关的蛋白质。用具有相同转录行为的已知基因将未知基因进行分类，来鉴定未知基因的功能，有的时候也被称为“牵连犯罪”。群分析已被用于分析许多有趣的生物过程下基因表达的轮廓，包括人类的伤口愈合（图 8-63）。

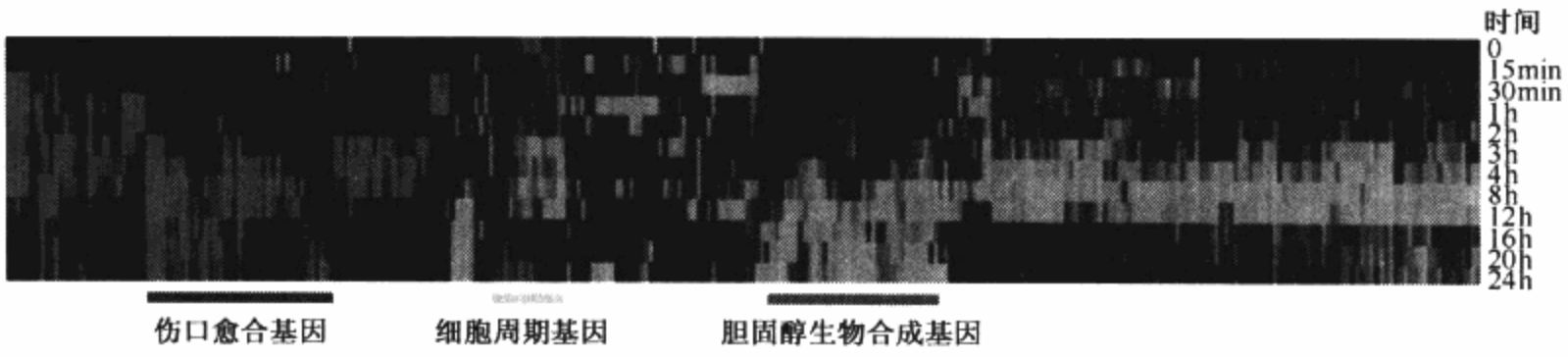


图 8-63　利用群分析鉴定协同调节的基因系列

属于同一群的基因可能与共同的细胞通路或进程相关。为了进行群分析，要从暴露在各种不同条件下的细胞样品中获得芯片数据，并将表现出表达模式协同变化的基因群放到一起。在这一实验中，人类纤维原细胞离开血清 48h，然后在零时刻向培养物中加入血清，并在不同的时间点上收集细胞做芯片分析。在芯片上进行分析的 8600 个基因中，只有超过 300 个对重新加入血清表现了 3 倍或更大的基因表达模式变化。这里，红色指示表达的增加，绿色指示表达的减少。在许多芯片实验的基础上，8600 个基因按相似的表达模式分成群。这一分析的结果表明，伤口愈合的相关基因响应血清而开启，而与细胞周期调节相关的基因和胆固醇生物合成的基因被关闭（摘自 M. B. Eisen et al., *Proc. Natl. Acad. Sci. USA* 95：14863～14868，1998）

目标突变可以揭示基因功能

尽管要在繁殖迅速的生物体中获得某一特异过程（如 DNA 复制和眼睛发育）的突变并非难事，但是要追踪这一突变到某一个特异改变的蛋白质是需要花费很长时间的。最近，重组 DNA 技术和基因组测序的暴增使得一种不同的遗传学研究方法成为可能。可以从一个特定的基因入手，在基因内部进行突变产生突变的细胞或生物体来分析基因的功能，而不是开始于一个随机产生的突变并利用它来鉴定基因和相应的蛋白质。由于这种新方法是从基因和蛋白质到突变体，颠倒了遗传学发现的传统方向，这种方法也被称为**反向遗传学**。

反向遗传学从一个克隆的基因，或从细胞中分离出来的具有有趣特性的蛋白质，或简单的基因组序列入手。如果起点是一个蛋白质，首先要鉴定编码它的基因，然后如果有必要的话，测定其核苷酸序列。在体外改变这段基因序列产生一个突变版本。这段改造的突变基因和一段合适的调节区域一起转入细胞中，它在细胞内整合到一条染色体上，永远成为细胞基因组的一部分。所有改造的细胞的后代都会含有突变基因。

如果用于基因转化的原始细胞是受精卵，如果突变不是致死性的，就可以获得含有突变

基因的完整多细胞生物体。在某些这种动物中，改变的基因会混合到生殖细胞中，产生生殖细胞突变，使得突变基因传到后代。

这种遗传学改造现在已经程序化地用于像果蝇和哺乳动物一样复杂的生物体上。从技术上讲，甚至人类也可以通过这种方法进行改造，然而目前即使以治疗为目的也没有施行这种程序，因为人们担心在这种个体中会发生不可预料的异常变化。

在本章的前面部分我们讨论过其他用于发现基因功能的方法，包括在其他生物体中搜索同源基因和测定基因何时何地表达。这类信息对于提示在突变的生物体中寻找哪些类型的突变特别有用。例如，一个只在成体肝脏中表达的基因，可能与降解毒素有关，但不太可能影响眼睛的发育。所有的这些方法都可以用于研究单个基因或试着对生物体每个基因的功能进行大范围的分析（一个处于萌芽阶段的领域——功能基因组学）。

可以按照要求产生含有突变基因的细胞和动物

我们已经知道搜索同源基因和分析基因表达模式可以提供关于基因功能的线索，但是它们不能揭示一个基因到底在细胞内做什么。遗传学提供了一种有用的方法来解决这一问题，因为缺少某一特殊基因的突变体可以快速揭示基因编码的蛋白质的功能。遗传工程可以特异地产生这些基因的敲除突变，下面我们将要看到。此外，还可以产生在异常高的水平上表达基因（过表达）、在错误的组织或错误的时间表达基因（错表达），或者以微小改变而产生显性表型的方式表达基因的突变体。为了促进这种基因功能研究，可以通过改造基因的编码序列和调节区域来改变蛋白质产物的功能特性、蛋白质的产量，或产生蛋白质的细胞类型。

可以通过各种方法将改造的基因导入细胞，一些方法将在第 9 章详细叙述。DNA 可以用玻璃的微量移液器微注射进入哺乳动物细胞，或者通过已经改造好可以携带外源基因的病毒导入细胞。在植物细胞中，基因常常通过一种叫做粒子轰击的技术导入细胞：将 DNA 样品涂到小的金粒上，然后用特制的枪逐个精确地射击，穿透细胞壁。电穿孔是用于介导 DNA 进入细菌或其他细胞的方法。在这种技术中，一个短暂的电击使得细胞膜具有短暂的通透性，允许外源的 DNA 进入细胞质。

现在我们就来看一下对这些突变细胞和生物体的研究如何解释各种生物学通路。

在细菌和一些低等真核生物中细胞中的正常基因可以直接被改造的突变基因代替

与高等真核生物（多细胞的二倍体）不同，细菌、酵母和黏菌一般以单倍体的单细胞形式存在。在这些生物体中，一个含有突变基因的人工导入的 DNA 分子，可以通过同源重组，以相对高的频率来替代单拷贝的正常基因，因此可以很容易地产生由突变基因代替了正常基因的细胞［图 8-64(A)］。通过这种方法，可以按照要求制造细胞，产生任何特异蛋白质或 RNA 分子代替正常形式的分子。如果突变基因是完全失活的，而且正常基因产物具有至关紧要的功能，细胞就会死掉；但是在这种情况下，可以用不严重突变的基因形式来代替正常基因，所以突变细胞可以存活下来而需要这种基因的进程是异常的。一般选择的突变体是可以产生温度敏感基因产物的个体，它的这种基因产物在某一温度下可以正常发挥功能，但当细胞转移到高一点儿或低一点儿的温度下就会失活。

低等真核生物的这种能够直接进行基因替代的能力，结合在这些单倍体生物体中的强有力的标准遗传分析方法，可以大体解释在这类细胞中的研究为什么会对了解真核生物中普遍

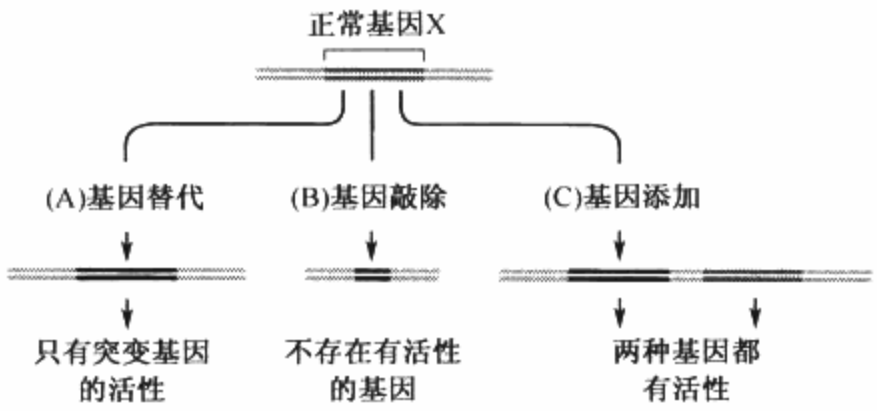


图 8-64　基因替代、基因敲除和基因添加

在遗传改造的生物中，可以通过许多方法改变一个正常基因。（A）正常基因（绿色）可以完全被基因的突变拷贝（红色）所代替，这一过程叫做基因替代。这种方法提供了不被正常基因干扰的突变基因活性的信息，因此就可以测定小而精细的突变的影响。（B）正常基因可以被完全失活，如在其中制造一大段缺失，这叫做基因敲除。（C）一个突变基因可以简单地添加到基因组上。在某些生物体中，这是最简单的一类遗传学改造。当引入的突变基因使正常基因的功能失效时，这种方法可以提供有用的信息

存在进程的细节是如此的重要。接下来我们将要看到，基因替代是可能的，但是在真核生物中更加困难，由于某些未完全理解的原因。

改造的基因可以用来在二倍体生物中产生特殊的显性失活突变

高等真核生物是二倍体的，如哺乳动物、果蝇、线虫，因此每个染色体有两个拷贝。并且被改造过的基因的转染一般导致基因添加而非基因替代：被改造的基因随机插入基因组中，因此细胞（或生物体）含有突变基因和正常基因拷贝。

由于在高等真核生物中进行基因添加要比基因替代容易得多，产生显性失活突变就很有用了，这种突变基因可以在细胞中消除其相应的正常基因的活性。一种巧妙的方法利用了两条互补核酸链的杂交反应特异性。通常地，双螺旋上特定部分的两条 DNA 链中只有一条可以被转录成 RNA，而且常常是特定基因的同一条链（图 6-14）。如果一个克隆的基因被改造了，使得与它相反的 DNA 链代替它被转录，就会产生与正常 RNA 转录物序列互补的反义 RNA 分子。当这种反义 RNA 大量合成时，就会与正常基因产生的“正义”RNA 杂交，从而抑制相应蛋白质的合成（图 8-65）。相似的方法包括化学或酶促合成短的反义核酸分子，然后将它们注射入（或用其他方法送入）细胞，阻断（尽管只是暂时的）相应蛋白质的产生。为了避免注入的核酸降解，用一种叫吗啉基 RNA 的稳定的 RNA 类似物，来代替普通 RNA。

当研究人员继续探究反义 RNA 策略时，他们有一个有趣的发现：一条反义 RNA 链能阻断基因表达，但是包含目标基因的正义链和反义链的双链 RNA，能够更有效地抑制目标基因的活性（图 7-107）。这种现象叫做链 RNA 干扰（RNAi），现在已经用于在一些生物体中检查基因的功能。

RNAi 技术已经广泛用于在线虫中研究基因的功能。当研究对象为线虫时，导入双链 RNA 是非常简单的：RNA 可以直接注射入动物的肠内，或者用表达目标基因双链 RNA 的大肠杆菌喂食线虫［图 8-66(A)］。这种 RNA 可以散布到线虫身体的全部，在不同的组织中抑制目标基因的表达。进一步来说，就像在图 7-107 中解释的那样，干扰可以高频率地遗传给被注射动物的后代。由于线虫的全基因组已经经过测序，利用 RNAi 可以用来指派所有线虫基因的功能。在一个研究中，研究人员们可以抑制线虫 3 号染色体上大约 2300 种预测基因

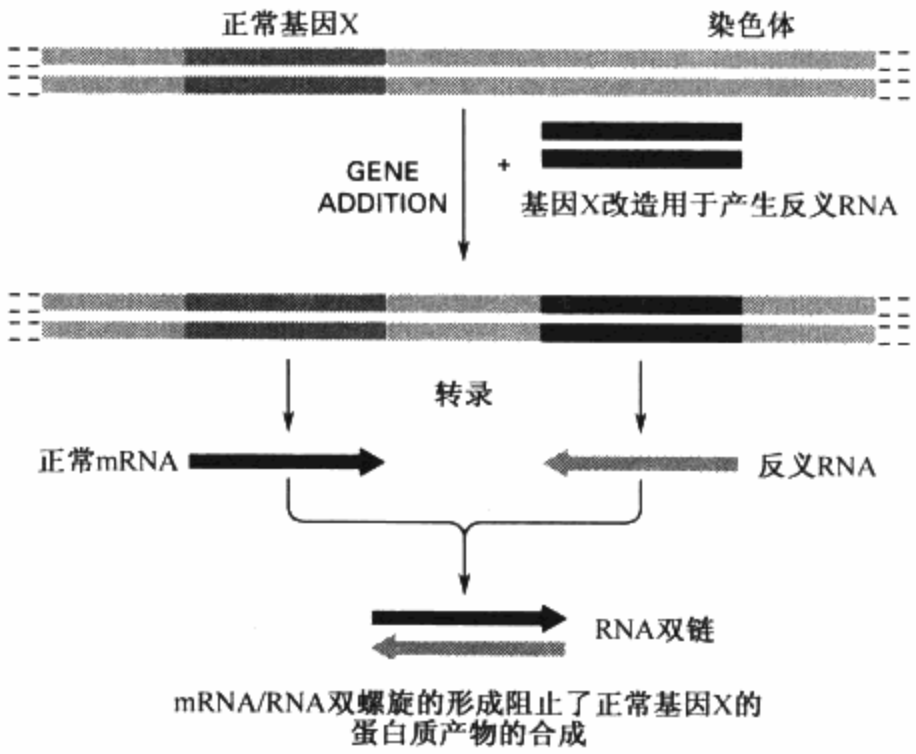


图 8-65　产生显性失活突变的反义 RNA 策略

与正常基因 X 产生的序列互补、用于产生反义 RNA 的改造基因，可以导致细胞内产生双链 RNA。如果产生了过量的反义 RNA，它就可以与基因 X 产生的正常 RNA 中的大多数杂交并使之失活。尽管将来用这种方法可能会使任何基因失活，但目前这一技术似乎只对某些基因管用，其他的则不行

的 96%。用这种方法，他们鉴定了与线虫胚胎的细胞分裂有关的 133 种基因［图 8-66(C)］。在这些基因当中只有 11 个以前通过实验确定了其功能。

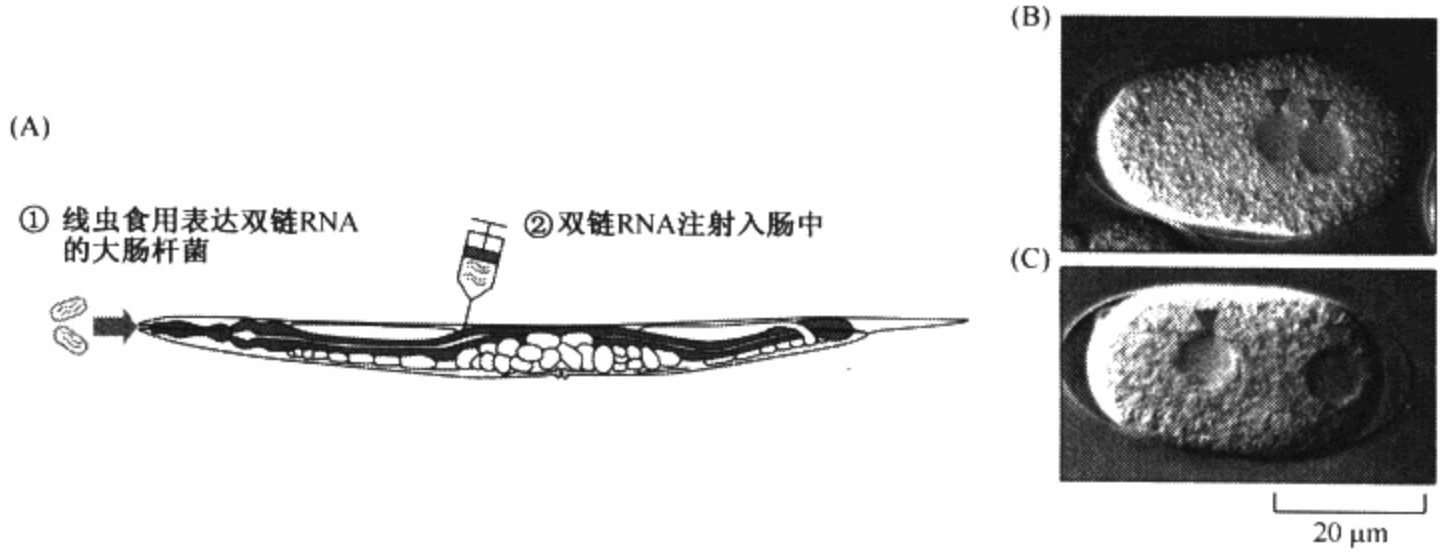


图 8-66　RNA 干扰引起的显性失活突变

（A）双链 RNA（dsRNA）可以通过以下方法导入线虫：①用表达目标基因双链 RNA 的大肠杆菌喂食线虫；②直接向肠中注入双链 RNA。（B）野生型线虫胚胎。（C）用 RNAi 失活与细胞分裂相关的一个基因的线虫胚胎。胚胎表现出未融合的卵和精子的细胞核的异常迁移［（B）和（C）摘自 P. Gonczy et al.，*Nature* 408：331～336，2000］

由于某些不知道的原因，RNA 干扰不能有效失活所有基因，而且有的时候干扰能抑制目标基因在一个组织的活性，却无法在另一个组织中抑制其活性。另一个可用来产生显性失活突变的方法利用了这样一个事实，即大多数蛋白质作为一个大的蛋白质复合体的一部分来发挥功能。通常只要包含一个无功能的组分，这些复合体就会被失活。因此，通过设计一个可以产生大量的无活性却能组装到复合体中的突变蛋白质的基因，就可能产生一种尽管存在正常蛋白质却使所有复合体都失活的细胞（图 8-67）。

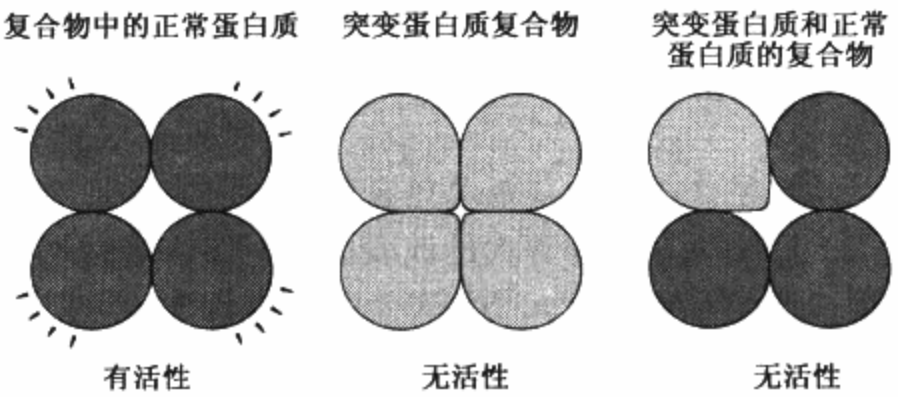


图 8-67　一种蛋白质的显性失活效应

在这里，一个基因经过改造，可以产生大量突变蛋白质，这种突变蛋白质可以阻止正常拷贝的同种蛋白质行使功能。在这个简单的例子中，正常蛋白质必须形成一个多亚基的复合物才有活性，突变蛋白质通过与正常蛋白质形成无活性的混合复合物来阻断其功能。通过这种方法，在基因组任何位置的一个单拷贝的突变基因，可以失活其他基因拷贝产生的正常产物

如果一个蛋白质是细胞（或生物体）生存所需的，一个显性失活突变体会死掉，就不能检测蛋白质的功能。为了避免这一问题，可以将突变基因与改造的控制基因结合，这种控制基因只能在命令时产生基因产物，如对温度的升高或某一特异信号分子的存在产生反应。含有这种由可诱导启动子控制的显性突变基因的细胞或生物体可以使特定的蛋白质在特定的时间失活，然后追踪这一变化产生的影响。可诱导的启动子也可以使特定组织中的基因开启或关闭，用来检测突变基因在生物体某一部分的效应。将来，产生显性失活突变来使特定基因失活的技术，可能广泛用于测定高等生物的蛋白质功能。

功能获得突变可为基因在细胞或生物体中发挥的功能提供线索

细胞可以通过改造表达显性失活的蛋白质，产生功能丧失的表型，用同样的方法改造细胞，通过功能获得突变也可以使它们表现一种异常的表型。这种突变可能赋予某一特定蛋白质一个新的活性，或者引起动物中正常活性的蛋白质在不合适的时间或错误的组织表达。不管是通过什么机制，功能获得突变可以在细胞、组织、生物体中产生新的表型。

功能获得突变一般是通过在细胞中过高的表达一个基因而产生的。实现这种过表达可以通过以下方法：将基因与强启动子序列结合后放在多拷贝的质粒中，将基因整合入基因组的多拷贝区。在每种情况下，基因都存在许多拷贝，而且每一拷贝都会指导产生大量 mRNA 分子。尽管必须小心地解释这种过表达对生物体表型的影响，但这种方法提供了对许多基因活性的宝贵认识。另一种功能获得突变中，突变蛋白质有正常的产量，但是活性远大于正常蛋白质。这种蛋白质常常在肿瘤中发现，它们已经被用于研究细胞中的信号转导途径（讨论见第 15 章）。

基因也可以在错误的时间或者生物体中错误的地点被表达，经常产生惊人的结果（图8-68）。这种错误表

图 8-68　影响爪蟾早期胚胎体轴发育的信号蛋白 Wnt 的异常表达

在这一实验中，将编码 Wnt 的 mRNA 注射入腹部的植物极，诱导产生了第二个体轴（摘自 S. Sokol et al., *Cell* 67: 741～752, 1992）

达通常是由基因自身的重改造产生的，这种重改造为改变它们的表达提供了需要的调节序列。

可以重新设计基因来产生有任何想要序列的蛋白质

研究基因及其编码蛋白质的作用时，并不总是希望进行诸如使细胞中充满大量的高活性的蛋白质，或者将一个基因的产物完全除掉等大的改变。有时候，稍稍改变蛋白质结构是很有用的，这样可以仔细研究蛋白质的哪一部分对功能是重要的。例如，可以通过改变活性位点的一个氨基酸来研究酶的活性。如此精细的改变基因和其蛋白质产物，需要特殊的技术。第一步通常是化学合成短 DNA 分子，这个分子含有部分改变的想要基因的核苷酸序列。合成的 DNA 寡核苷酸，在允许 DNA 链非完全配对的条件下，与含有待改造的 DNA 序列的单链质粒 DNA 杂交（图 8-69）。合成的寡核苷酸可以作为引物，通过 DNA 聚合酶进行 DNA 合成，从而产生在一条链上具有已改造序列的 DNA 双螺旋。转染之后，就可以得到带有完全改造的基因序列的质粒。然后这种正确的 DNA 就可以插入到表达载体中，在合适的细胞中表达重新设计的蛋白质，用于研究其功能。通过这种方法改变蛋白质中选定的氨基酸，即**定点诱变**，可以准确地测定多肽链的哪些部分对诸如蛋白质折叠、与其他蛋白质的相互作用和酶催化等过程是重要的。

改造的基因可以很容易地插入许多动物的生殖系

当改造一个生物体使其表达变异基因的时候，理想的状况是能够用变异的基因来替代正常基因，以便在没有正常蛋白质存在的情况下分析突变蛋白质的功能。像前面所讨论的，这一过程可以在一些单倍体的单细胞生物中很容易的完成。我们将要在下面的部分看到，已经发展出了更加复杂的流程，在小鼠中产生这种类型的基因替代。但是，外源 DNA 可以非常简单地随机整合到许多动物基因组中。例如，在哺乳动物中，导入细胞的线性 DNA 片段会通过细胞内的酶进行快速的末端对末端连接，形成长的串联排列，这种阵列通常在明显的随机位点整合入染色体。哺乳动物受精卵在这一方面的表现和其他哺乳动物细胞一样。注射了 200 个线性 DNA 分子拷贝的小鼠卵细胞常常发展成在很多细胞中含有注射基因的小鼠，这些注射基因的拷贝形成串联排列整合到某一染色体的一个随机位点。如果更改的染色体出现在生殖细胞系中（卵细胞或精子），小鼠就会将这些外源基因传给子代。

通过基因插入、基因删除或基因替代而永久性改变的生物称为转基因生物，任何添加的外源或改造基因叫做转基因。当正常基因仍然存在时，只有显性的改变会在表型分析中表现出来。尽管如此，具有插入基因的转基因动物为认识哺乳动物基因如何调节和某些特异变化的基因（癌基因）如何引起癌症提供了重要依据。

产生转基因果蝇也是可能的，在这种果蝇中，某一基因的单拷贝随机插入到果蝇基因组中。在这种情况下，DNA 片段首先要插入到果蝇转座子 P 因子的两个末端序列之间。在 P 因子转座酶存在时，末端序列使得 P 因子可以整合到果蝇染色体中。为了得到转基因果蝇，适当改造的 DNA 片段和含有转座酶基因的独立质粒一起，注入到极早期的果蝇胚胎中。这一过程完成后，注入的基因在转座之后，常常以单拷贝的方式进入生殖系。

基因打靶可以产生缺失特异基因的转基因小鼠

如果一个带有突变小鼠基因的 DNA 分子转入小鼠细胞中，它通常随机地插入染色体，但是大概在 1000 次里面有一次它会通过同源重组代替两个拷贝的正常基因中的一个。通过研究这些罕见的“基因打靶”现象，小鼠中的任何特异的基因都可以通过直接的基因替代来改变或失活。如果是将感兴趣的基因失活，产生的动物就叫做“敲除”小鼠。

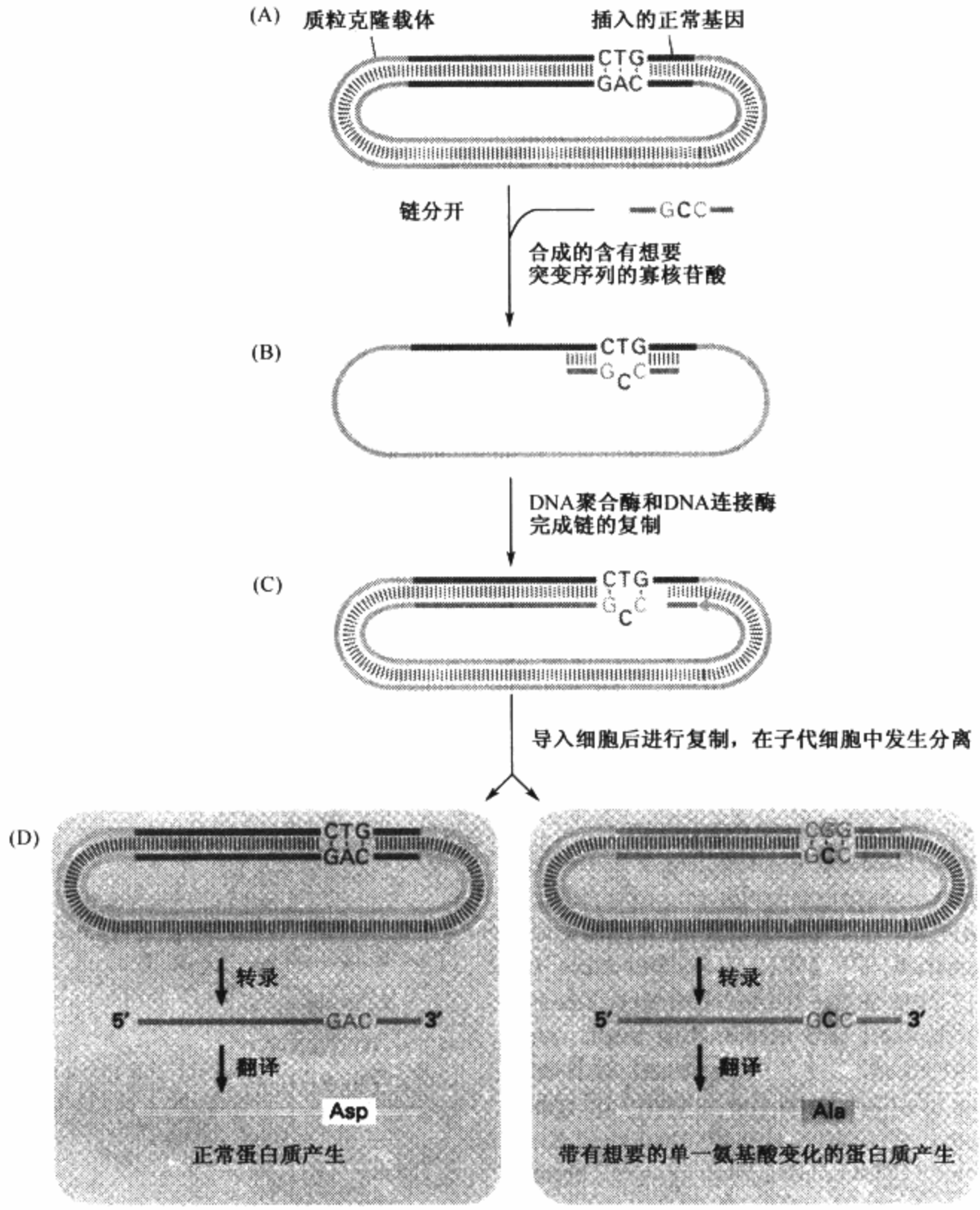


图 8-69　在定点诱变方法中，利用合成的寡核苷酸来改变基因的蛋白质编码区

（A）含有插入基因的重组质粒被分成两条 DNA 链。在低于 DNA 完全杂交的条件下，向 DNA 单链中加入合成的 DNA 寡核苷酸引物，这一引物符合基因序列的一部分但是在预定位点改变了一个核苷酸。（B）引物与 DNA 杂交，形成有一个核苷酸错配的核酸对。（C）重组质粒在体外从引物开始进行 DNA 合成形成双链，由连接酶封口。（D）双链 DNA 导入细胞，进行复制。用一条链作为模板的复制产生了一个正常的 DNA 分子，但是利用另一条链（含有引物）作为模板的复制产生了带有想要的突变的 DNA 分子。只有一半的子代含有一个带有想要的突变基因的质粒。但是含有突变基因的子代细胞可以从其他细胞中鉴定、分离出来，培养产生带有突变基因的纯细胞系。这里只显示了可以改造的许多变化中的一种。利用合适序列的寡核苷酸，可以一次进行多于一个的氨基酸替代，或者插入或删除一个或多个氨基酸。尽管没有在这个图中表示出来，也可以利用合适的寡核苷酸和 PCR（代替质粒复制）扩增突变基因，来产生定点突变

这一技术的流程是这样的：第一步，一个含有想要的突变基因 DNA 片段（或设计的用来打断目标基因的 DNA 片段）插入到载体中，然后导入从小鼠干细胞衍生出的特异胚胎细胞系，即**胚胎干细胞**或 ES 细胞，这种细胞可以在细胞培养物中生长而且能够产生许多不同

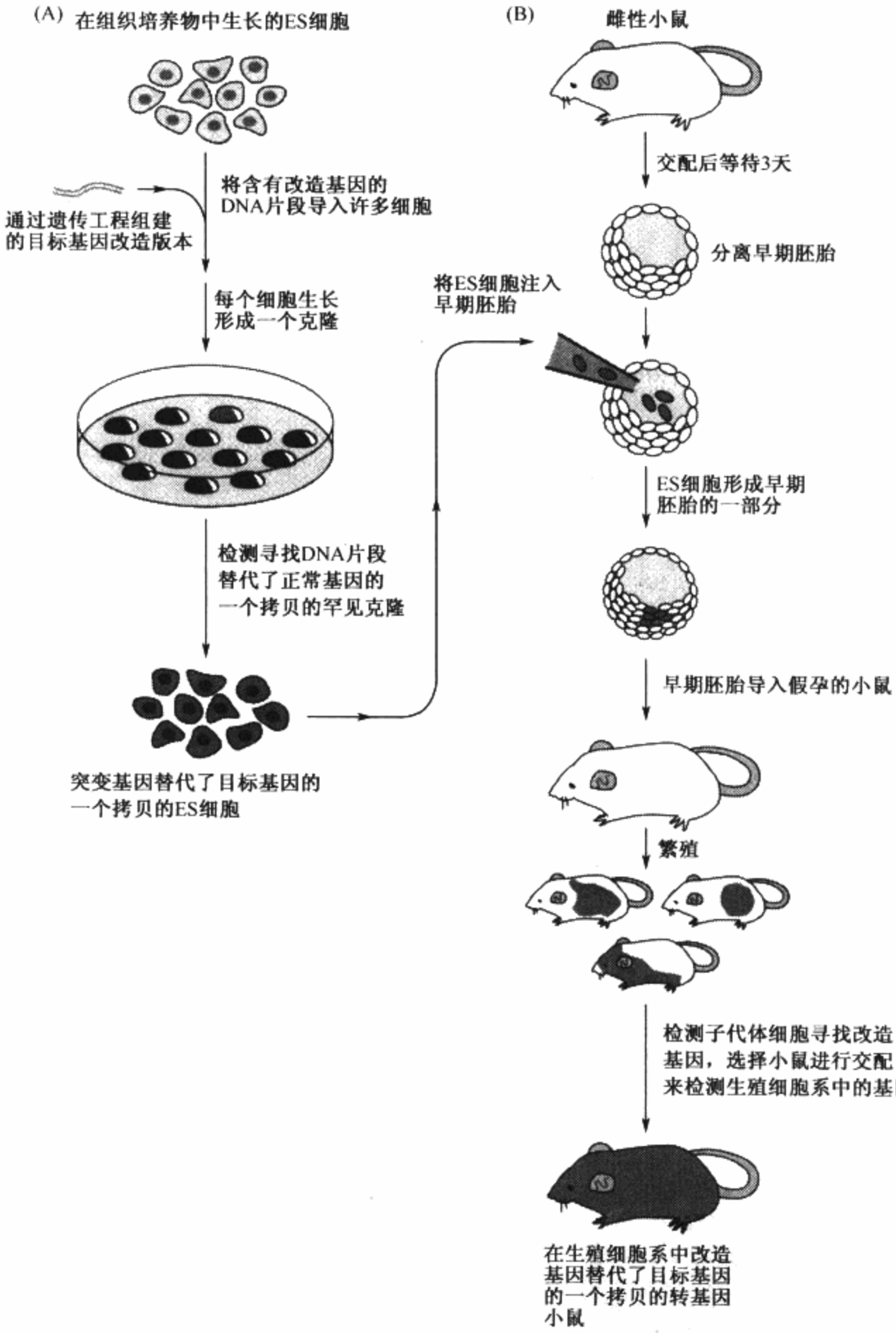


图 8-70　用于在小鼠中产生基因替代的流程总结

（A）基因的改造版本导入培养的 ES 细胞（胚胎干细胞）。只有很罕见的 ES 细胞会通过同源重组将相应的正常基因替代为改造的基因。尽管这一流程通常是费力的，这些罕见的细胞可以被鉴定并培养产生许多后代，每个后代都带有在两个相应正常基因的位点之一上的改造基因。（B）改造的 ES 细胞注入很早期的小鼠胚胎，这些细胞与生长的胚胎混合，通过这种胚胎产生的小鼠会含有一些携带改造基因的体细胞（用橙色表示）。一些小鼠还会含有携带改造基因的生殖细胞系。当与正常小鼠交配时，这些小鼠的一些后代会在所有的细胞中带有改造基因。如果接着用这种小鼠交配（没有在图中表示），一些后代会在所有细胞中都含有两个改造的基因（一条染色体上各一个）。

如果最初的基因改造使得基因功能完全失活，这些小鼠就是敲除小鼠。如果这些小鼠缺失的基因在发育中具有功能，它们通常会因为特异的缺陷在成年之前死掉。仔细分析这些缺陷可以帮助解释缺失基因的正常功能

组织类型的细胞。细胞增殖一段时间后，分离出可能通过同源重组发生了基因替代的罕见细胞克隆。从中得到的正确克隆通过 PCR 或者 Southern 印迹来鉴定：它们含有重组 DNA 序列，其上的插入片段已经替代了正常基因的一个拷贝的全部或部分。第二步，用微量吸量管吸取已鉴定克隆中的个体细胞，并将其注入早期的小鼠胚胎。转染的胚胎干细胞与宿主胚胎的细胞一同产生外观正常的小鼠；这种动物的大部分，在有利的情况下包括生殖细胞系，通常来自人工改造的干细胞（图 8-70）。

在生殖细胞系中含有转基因的小鼠交配后产生一个雌性动物和一个雄性动物，它们每个都是基因替代的杂合子（也就是说，它们具有一个拷贝的正常基因和一个拷贝的突变基因）。当接下来用这两个小鼠进行交配时，它们的子代有 1/4 会是改造基因的纯合子。研究这些纯合子可以检查在相应正常基因不存在时改造基因的功能或者说是消除一个基因活性产生的效应。

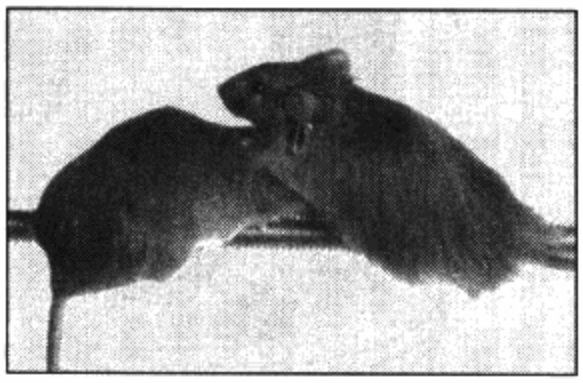


图 8-71　具有纤维原细胞生长因子 5（FGF5）缺陷的改造小鼠

FGF5 是毛发形成的负调节因子。在缺失 FGF5 的小鼠（右边）中，毛发比同窝杂合子个体（左边）要长。已经产生了具有和人类各种病症类似表型的转基因小鼠，这些病症包括老年痴呆症、动脉硬化症、糖尿病、囊肿性纤维化和某些癌症。对它们的研究可能产生更有效的治疗方式（经 Gail Martin 许可，摘自 J. M. Hebert et al., *Cell* 78: 1017～1025, 1994）

制备已知正常基因缺失的转基因小鼠是一个很大的突破，这种技术已经用于仔细研究大量的哺乳动物的基因功能（图 8-71）。相关的技术可用于产生条件突变体，可以在特异组织和发育过程中的特定时间来破坏选定基因。这种策略利用了位点特异的重组系统，在特殊位置或特殊时间切除并失活目标基因。这种重组系统最常见的叫做 **Cre/lox**，广泛用于小鼠和植物中的基因替代（图 5-82）。这种系统中，ES 细胞中的目标基因被侧翼夹有一对短 DNA 序列的具有完整功能的基因版本取代，这种侧翼序列叫做 lox 位点，它可被 Cre 重组酶识别。然后它们与在可诱导启动子的控制下表达 Cre 重组酶基因的转基因小鼠交配。在 Cre 开启的特异细胞或组织中，重组酶催化了 lox 序列间的重组，切除了目标基因，消除了它的作用。相似的重组系统用于在果蝇中产生条件突变体（图 21-48）。

转基因植物对细胞生物学和农业都是重要的

当植物受到损伤时，它常常可以通过已分化的成熟细胞的“去分化”、增殖，重新分化成其他细胞类型的过程来进行自身的修复。在某些情况下，去分化的细胞甚至可以形成顶端分生组织，顶端分生组织能够产生一个包括配子的完整新植物。这种非凡的可塑性可以用于在培养物中生长的细胞中产生转基因植物。

当一块植物组织在含有营养物和适当的生长调节因子的灭菌培养基中进行培养时，许多细胞以无组织的方式不确定的增殖，产生未分化的细胞群，叫做愈伤组织。如果仔细地调控营养物和生长调节因子，就可以在愈伤组织中先诱导形成新芽，然后是根的顶端分生组织，在许多物种中，可以产生完整的新植株。

愈伤组织培养物也可以被分成单个的细胞，这些细胞会生长和分裂形成悬浮培养物。在许多植物中，包括烟草、矮牵牛花、胡萝卜、马铃薯和拟南芥，来自这种悬浮培养物的单个细胞可以长成一个小块（一个克隆），可以从中重新产生完整的植株。这种具有产生生物体所有部分的能力的细胞，被认为是**全能的**。类似于从培养的胚胎干细胞的基因操作中得到转基因小鼠的过程，转基因植物也可以由培养中的单个全能的植物细胞转染了 DNA 之后产生（图 8-72）。

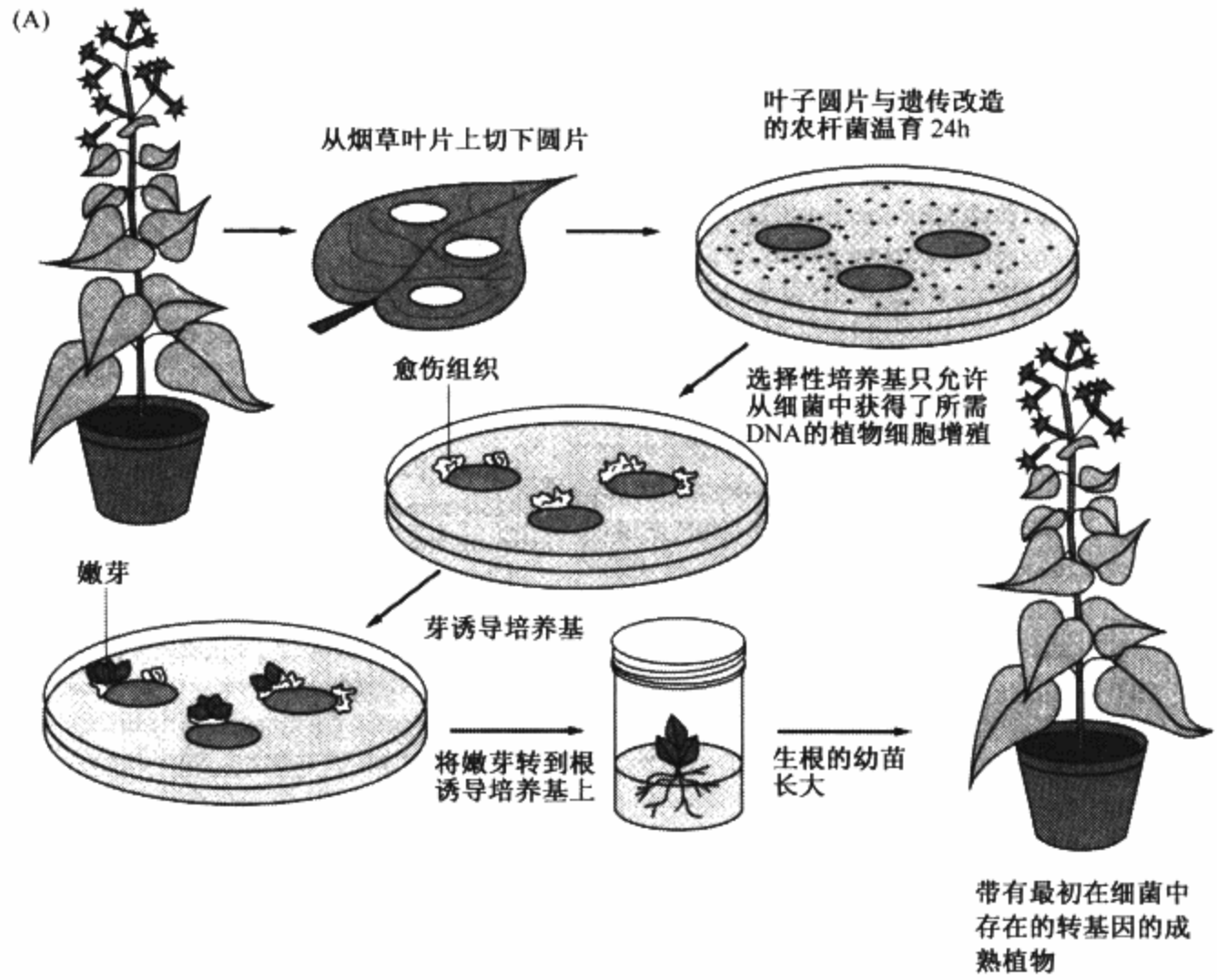


图 8-72　产生转基因植物的流程

（A）流程概述。从叶子上切下一个小圆片，和带有重组质粒的农杆菌共培养，重组质粒上有可选择的标记和想要的转基因。圆片边缘的受伤细胞释放物质吸引农杆菌，使它们将 DNA 注入这些细胞。只有那些吸收了合适 DNA 并且表达了选择性标记基因的植物细胞可以存活、增殖，形成愈伤组织。调节供给愈伤组织的生长因子，诱导它先形成新芽，然后形成根，并长成带有转基因的成熟植物。（B）重组质粒的制备及其向植物细胞的转化。一个农杆菌质粒通常带有 T-DNA 序列，可以用选择性标记（如卡那霉素抗性基因）和想要的转基因替代 25 个核苷酸对长的 T-DNA 之间的序列来进行改造。当农杆菌识别了一个植物细胞，它会利用一种转化质粒 T-DNA 序列的特殊机制，将一条携带这些序列的 DNA 链有效地传给植物细胞

转基因植物的产生极大地促进了植物细胞生物学的发展。例如，在分离生长调节因子和分析植物中形态发生和基因表达的机制的研究过程中，它都起到了重要作用。它也为农业向着既有利于农民又有利于消费者的方向发展提供了可能。它使得很多改造成为可能。例如，改变种子中的脂类、淀粉、蛋白质的含量，赋予植物虫害和病毒的抗性，产生可以忍受极端环境（如盐碱地和干旱土壤）的改造植物。

理解动物发育的许多重大进展来自于对果蝇和线虫的研究，对这两种动物既可以进行遗传学分析，又可以进行实验操作。相比之下，过去植物发育生物学的进展相对缓慢。许多可以用于遗传学分析的植物，如玉米和番茄，具有很长的生活周期和很大的基因组，使得经典的和分子的遗传分析都很费时间。后来，越来越多的人注意到一种生长迅速的野草、常见的十字花科植物——拟南芥，它具有很多作为“模式植物”的优点（图 1-46 和图 21-107）。相对较小的拟南芥基因组是第一个被完整测序的植物基因组。

大量收集标记的敲除突变体为检测生物体中每个基因的功能提供了工具

为了得到许多模式生物（包括酵母、线虫、果蝇、拟南芥和小鼠）的全面突变文库，研究者们正在尝试广泛的合作。其终极目标是得到一个突变菌株的集合，这些突变中生物体的每个基因或者被系统的删除，或者被改造成可以有条件的中断。这种类型的集合会为在基因组水平上研究基因功能提供宝贵的工具。在某些情况下，集合中的每个突变体都会显示一个独特的分子标签——一段特异的 DNA 序列用来快速常规地识别改造基因。

由于酵母具有同源重组的倾向，在酵母中产生一套 6000 个突变体且每个突变体只缺失一个基因的任务还算简单。为每个基因准备一个“删除盒子”，盒子由一个特殊的 DNA 分子构成，这种分子含有一个选择性标记，选择标记两侧是和目标基因每个末端序列一致的 50 个核苷酸。此外，一个特殊的“条码”序列也在这个分子中，以便于之后对突变菌株的快速识别（图 8-73）。然后，这种基因敲除突变的大量混合物在不同的选择性条件下（如营养物缺乏、

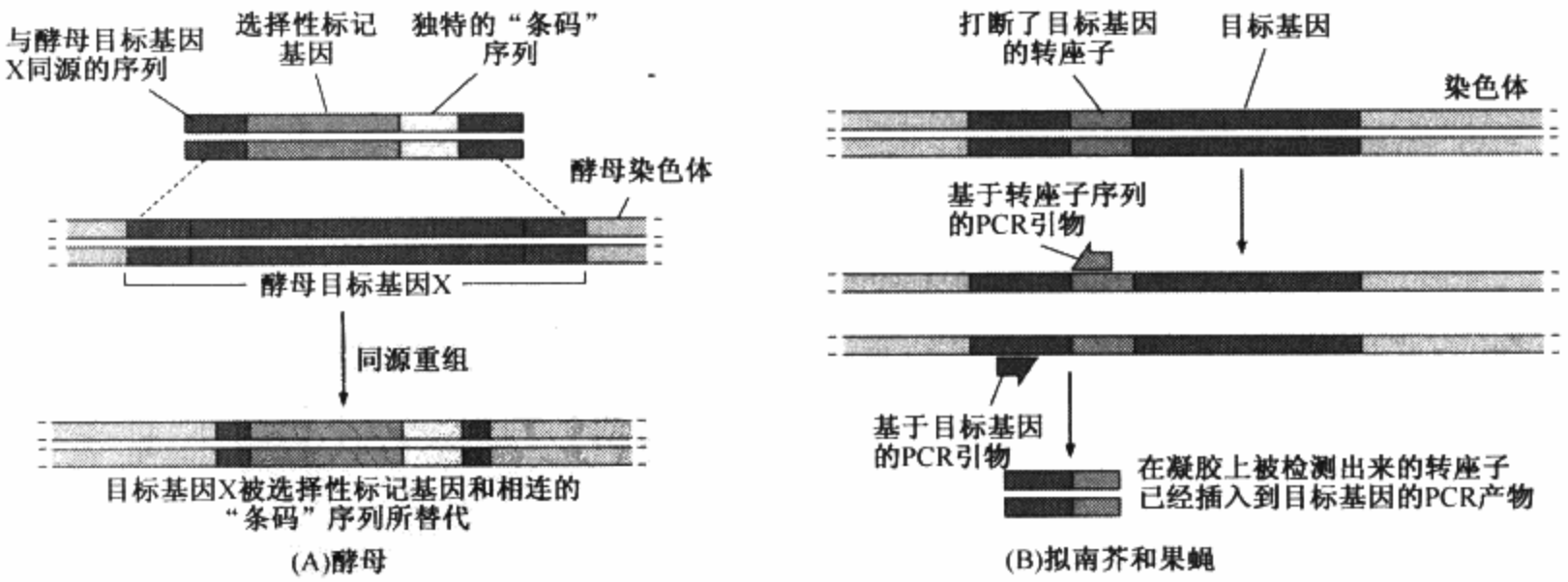


图 8-73　得到突变生物体的集合

（A）在酵母中使用的删除盒子含有目标基因 X 的每个末端的同源序列（红色）、一个选择性标记（蓝色）、一个长约 20 个核苷酸的独特的“条码”序列（绿色）。这一 DNA 导入酵母，可以通过同源重组替代目标基因。（B）类似的方法可以在拟南芥和果蝇中用于制备标记的敲除突变。在这种情况下，利用转座子偶然插入到目标基因的过程来产生突变体。收集得到的生物体的全 DNA，利用与转座子和目标基因结合的 PCR 引物来快速筛选打断了感兴趣基因的突变体。只有转座子插入到目标基因的 PCR 产物才会在凝胶上被检测出来

温度变化、存在不同药物）生长，存活的细胞可以通过它们独特的序列标签进行识别。通过评价混合物中每个突变体存活的好坏，可以初步估计哪些基因对不同条件下的生长是必需的、有用的和不相关的。

这种从酵母突变体中得到信息的方法，其挑战在于在突变表型的基础上来推断基因的功能和生物学作用。一些缺陷，例如，不能在组氨酸缺乏的条件下生存，直接指出了野生型基因的功能，其他的关系可能不会这么明显。对冷的突然敏感可能暗示了某一特异基因在酵母细胞中的功能。这种问题在比酵母更复杂的生物体中更难回答。例如，小鼠中某一基因功能的缺失，在不同的发育阶段可能影响许多不同的组织类型，而其他基因缺失却可能没有明显的效应。充分鉴定小鼠的突变表型常常需要彻底的检查，同时结合广泛的有关小鼠的解剖学、组织学、病理学、生理学以及复杂行为的知识。

但是，通过检查突变体文库产生的见解将会是伟大的。例如，对 *Mycoplasma genitalium*（支原体）这种已知具有最小基因组生物的大量突变体的研究，已经鉴定出细胞生活所必需的最少的基因。对突变库的分析表明，支原体中的 480 个蛋白质编码基因中有 265～350 个是在实验室条件下生长所需要的。将近 100 个这种必需基因是功能未知的，这说明还有惊人数目的作为细胞生活基础的基本分子机制等待着被发现。

小结

遗传学和遗传工程为在细胞和生物体中研究基因功能提供了有力的工具。在经典的遗传学方法中，随机诱变后通过筛选来鉴定在某一特定生物进程中具有缺陷的突变体。这些突变体接着用于定位和研究对这一进程有作用的基因。

也可以提供反义遗传技术来确定基因功能。DNA 改造的方法可以用于突变任何基因并将其重新插入细胞的染色体上，以使其成为基因组永久的一部分。如果用于这种基因转化的细胞是受精卵（对动物而言）或者一个培养中的全能植物细胞，就可以产生表达突变基因并可将其传给后代的转基因生物。对细胞生物学而言特别重要的是，以高度特异的方式改变细胞和生物体的能力，可以洞悉单个的蛋白质或 RNA 分子的有计划改变的效应。

许多这样的方法经过扩展，用来在基因组水平上研究基因功能。例如，DNA 芯片等技术可以用于同时监测数千种基因的表达，提供了在复杂细胞进程之下的详细而又全面的基因表达快照。生物体中的每个基因被系统地删除或打断而产生的突变体的文库，为探索每个基因在使生命发生的精细分子协作过程中的功能提供了宝贵工具。

参考文献

一般文献

Ausubel FM, Brent R, Kingston RE et al. (eds) (1999) Short Protocols in Molecular Biology, 4th edn. New York: Wiley.

Brown TA (1999) Genomes. New York: Wiley-Liss.

Spector DL, Goldman RD, Leinwand LA (eds) (1998) Cells: A Laboratory Manual. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.

Watson JD, Gilman M, Witkowski J & Zoller M (1992) Recombinant DNA, 2nd edn. New York: WH Freeman.

Isolating Cells and Growing Them in Culture

Cohen S, Chang A, Boyer H & Helling R (1973) Construction of biologically functional bacterial plasmids *in vitro*. *Proc. Natl. Acad. Sci. USA* 70, 3240–3244.

Emmert-Buck, MR, Bonner RF, Smith PD et al. (1996) Laser capture microdissection. *Science* 274, 998–1001.

Freshney RI (2000) Culture of Animal Cells: A Manual of Basic Technique, 4th edn. New York: Wiley.

Ham RG (1965) Clonal growth of mammalian cells in a chemically defined,

synthetic medium. *Proc. Natl. Acad. Sci. USA* 53, 288–293.

Harlow E & Lane D (1999) Using Antibodies: A Laboratory Manual. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.

Herzenberg LA, Sweet RG & Herzenberg LA (1976) Fluorescence-activated cell sorting. *Sci. Am.* 234(3), 108–116.

Jackson D, Symons R & Berg P (1972) Biochemical method for inserting new genetic information into DNA of simian virus 40: circular SV40 DNA molecules containing lambda phage genes and the galactose operon of *Escherichia coli*. *Proc. Natl. Acad. Sci. USA* 69, 2904–2909.

Levi-Montalcini R (1987) The nerve growth factor thirty-five years later. *Science* 237, 1154–1162.

Milstein C (1980) Monoclonal antibodies. *Sci. Am.* 243(4), 66–74.

Fractionation of Cells

de Duve C & Beaufay H (1981) A short history of tissue fractionation. *J. Cell Biol.* 91, 293s–299s.

Laemmli UK (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* 227, 680–685.

Nirenberg MW & Matthaei JH (1961) The dependence of cell-free protein synthesis in *E. coli* on naturally occurring or synthetic polyribonucleotides. *Proc. Natl. Acad. Sci. USA* 47, 1588–1602.

O'Farrell PH (1975) High-resolution two-dimensional electrophoresis of proteins. *J. Biol. Chem.* 250, 4007–4021.

Palade G (1975) Intracellular aspects of the process of protein synthesis. *Science* 189, 347–358.

Pandey A & Mann M (2000) Proteomics to study genes and genomes. *Nature* 405, 837–846.

Scopes RK & Cantor CR (1993) Protein Purification: Principles and Practice, 3rd edn. New York: Springer-Verlag.

Yates JR (1998) Mass spectrometry and the age of the proteome. *J. Mass. Spectrom.* 33, 1–19.

Isolating, Cloning, and Sequencing DNA

Adams MD, Celniker SE, Holt RA et al. (2000) The genome sequence of *Drosophila melanogaster*. *Science* 287, 2185–2195.

Alwine JC, Kemp DJ & Stark GR (1977) Method for detection of specific RNAs in agarose gels by transfer to diabenzyloxymethyl-paper and hybridization with DNA probes. *Proc. Natl. Acad. Sci. USA* 74, 5350–5354.

Blattner FR, Plunkett G, Bloch CA et al. (1997) The complete genome sequence of *Escherichia coli* K-12. *Science* 277, 1453–1474.

Hunkapiller T, Kaiser RJ, Koop BK & Hood L Large-scale and automated DNA sequence determination. *Science* 254, 59–67.

International Human Genome Sequencing Consortium (2000) Initial sequencing and analysis of the human genome. *Nature* 409, 860–921.

Maniatis T et al. (1978) The isolation of structural genes from libraries of eucaryotic DNA. *Cell* 15, 687–701.

Saiki RK, Gelfand DH, Stoffel S et al. (1988) Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science* 239, 487–491.

Sambrook J, Russell D (2001) Molecular Cloning: A Laboratory Manual, 3rd edn. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.

Sanger F, Nicklen S & Coulson AR (1977) DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA* 74, 5463–5467.

Southern EM (1975) Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J. Mol. Biol.* 98, 503–517.

The *Arabidopsis* Genome Initiative (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* 408, 796–815.

The *C. elegans* Sequencing Consortium (1998) Genome sequence of the nematode *C. elegans*: a platform for investigating biology. *Science* 282, 2012–2018.

Venter JC, Adams MA, Myers EW et al. (2000) The sequence of the human genome. *Science* 291, 1304–1351.

Analyzing Protein Structure and Function

Bamdad C (1997) Surface plasmon resonance for measurements of biological interest, in Current Protocols in Molecular Biology (FM Ausubel, R Brent, RE Kingston et al. eds), pp 20.4.1–20.4.12. New York: Wiley.

Branden C & Tooze J (1999) Introduction to Protein Structure, 2nd edn. New York: Garland Publishing.

Fields S & Song O (1989) A novel genetic system to detect protein–protein interactions. *Nature* 340, 245–246.

Kendrew JC (1961) The three-dimensional structure of a protein molecule. *Sci. Am.* 205(6), 96–111.

MacBeath G & Schreiber SL (2000) Printing proteins as microarrays for high-throughput function determination. *Science* 289, 1760–1763.

Miyawaki A, Tsien RY (2000) Monitoring protein conformations and interactions by fluorescence resonance energy transfer between mutants of green fluorescent protein. *Methods Enzymol.* 327, 472–500.

Sali A & Kuriyan J (1999) Challenges at the frontiers of structural biology. *Trends Genet.* 15, M20–M24.

Wuthrich K (1989) Protein structure determination in solution by nuclear magnetic resonance spectroscopy. *Science* 243, 45–50.

Studying Gene Expression and Function

Botstein D, White RL, Skolnick M & Davis RW (1980) Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *Am. J. Hum. Genet.* 32, 314–331.

Brent R (2000) Genomic Biology. *Cell* 100, 169–182.

Capecchi MR (2001) Generating Mice with Targeted Mutations. *Nat. Med.* 7, 1086-1090.

Coelho PS, Kumar A & Snyder M (2000) Genome-wide mutant collections: toolboxes for functional genomics. *Curr. Opin. Microbiol.* 3, 309–315.

DeRisi JL, Iyer VR & Brown PO (1997) Exploring the metabolic and genetic control of gene expression on a genomic scale. *Science* 278, 680–686.

Eisenberg D, Marcotte EM, Xenarios I & Yeates TO (2000) Protein function in the post-genomic era. *Nature* 415, 823–826.

Enright AJ, Illiopoulos I, Kyrpides NC & Ouzounis CA (1999) Protein interaction maps for complete genomes based on gene fusion events. *Nature* 402, 86–90.

Hartwell LH, Hood L, Goldberg ML et al. (2000) Genetics: From Genes to Genomes. Boston: McGraw-Hill.

Lockhart DJ & Winzeler EA (2000) Genomics, gene expression and DNA arrays. *Nature* 405, 827–836, 2000.

Nusslein-Volhard C & Weischaus E (1980) Mutations affecting segment number and polarity in *Drosophila*. *Nature* 287, 795–801.

Palmiter RD & Brinster RL (1986) Germ line transformation of mice. *Annu. Rev. Genet.* 20, 465–499.

Rubin GM & Sprading AC (1982) Genetic transformation of *Drosophila* with transposable element vectors. *Science* 218, 348–353.

Tabara H, Grishok A & Mello CC (1998) RNAi in *C. elegans*: soaking in the genome sequence. *Science* 282, 430–431.

Weigel D & Glazebrook J (2001) *Arabidopsis*: A Laboratory Manual. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.

9 细胞显像

显微镜下观察细胞的结构
观察活细胞中的分子

细胞虽小，但很复杂。观察细胞的结构或揭示其分子组成都是不容易的，而阐明其不同组分的功能则更为困难。我们对细胞的了解依赖于我们所使用的工具，细胞生物学的主要进展常常是引进新技术的结果。因此，要理解现代的细胞生物学，我们必须知道其研究方法。

本章简要地回顾了用于研究细胞的一些主要显微术。理解细胞的组织结构是理解细胞如何发挥功能的必不可少的前提条件。光学显微术将是我们的起点，因为细胞生物学是从光学显微镜开始的，而且，现在它仍然是一项必不可少的工具。近年来，光学显微术变得尤为重要，这主要是因为单个细胞组分的特异标记方法和显像方法及其三维结构的重构方法的发展。光学显微术的主要优点是光相对而言没有破坏性。通过将特异细胞组分加上荧光标签，例如绿色荧光蛋白（GFP），我们因此得以观察它们在活细胞或有机体中的运动和相互作用。

光学显微术受限于它所能揭示的细节的精细程度。与使用可见光的显微镜相比，使用其他类型射线的显微镜——特别是电子显微镜——能解析小得多的结构。其代价是，电子显微术的样本制备要复杂得多，而且难以确定我们所见的影像是否能精确反映所检测标本的真实结构。然而，现在的电子显微术通过快速冰冻真实地保留结构还是可能的。计算机图像分析技术可用于从多重视图重构三维物体。连同这些方法，我们可将显微术的分辨率和使用范围扩展到能显示单个大分子的结构。

尽管方法极为重要，但是我们用这些方法所得到的发现使得它们更为有趣。因此，本章只是用作参考章节，应与本书的其他章节一起阅读，而不仅仅是介绍这些方法。

显微镜下观察细胞的结构

一个典型的动物细胞直径为 10～20μm，这大小相当于肉眼可见的最小颗粒的 1/5。19 世纪早期性能优良的光学显微镜得到应用，人们发现所有的植物和动物组织都是细胞个体的集合。这个由施莱登和施旺在 1838 年提出的发现，标志着细胞生物学的正式诞生。

动物细胞不仅仅很微小，而且是无色和透明的。因此，在 19 世纪晚期，随着一系列的能区别出细胞内部特征的染色方法的发展，细胞主要内部特征才得以发现。同样地，在 1940 年早期电子显微镜这种更有力的工具的引入，也需要依赖于细胞保存和染色新技术的发展，才能展现出细胞内部的细微结构。时至今日，显微技术既依赖于显微镜本身的性能，也依赖于样品制备的方法。因此，接下来的讨论中，我们从光学显微镜开始，介绍显微设备和样品制备方法。

图 9-1 通过从拇指到一簇原子的一系列图像阐明了尺寸大小的系列变化。每张照片比前一张的放大比率要大 10 倍。肉眼能看见前两幅图里的图像，光学显微镜分辨率能到第 4 幅图像左右，电子显微镜能到大概第 7 幅图。表 9-1 中列出了光学显微镜发展的里程碑。图 9-2 显示了多种细胞和亚细胞结构的尺寸以及不同类型显微镜观测尺寸的范围。

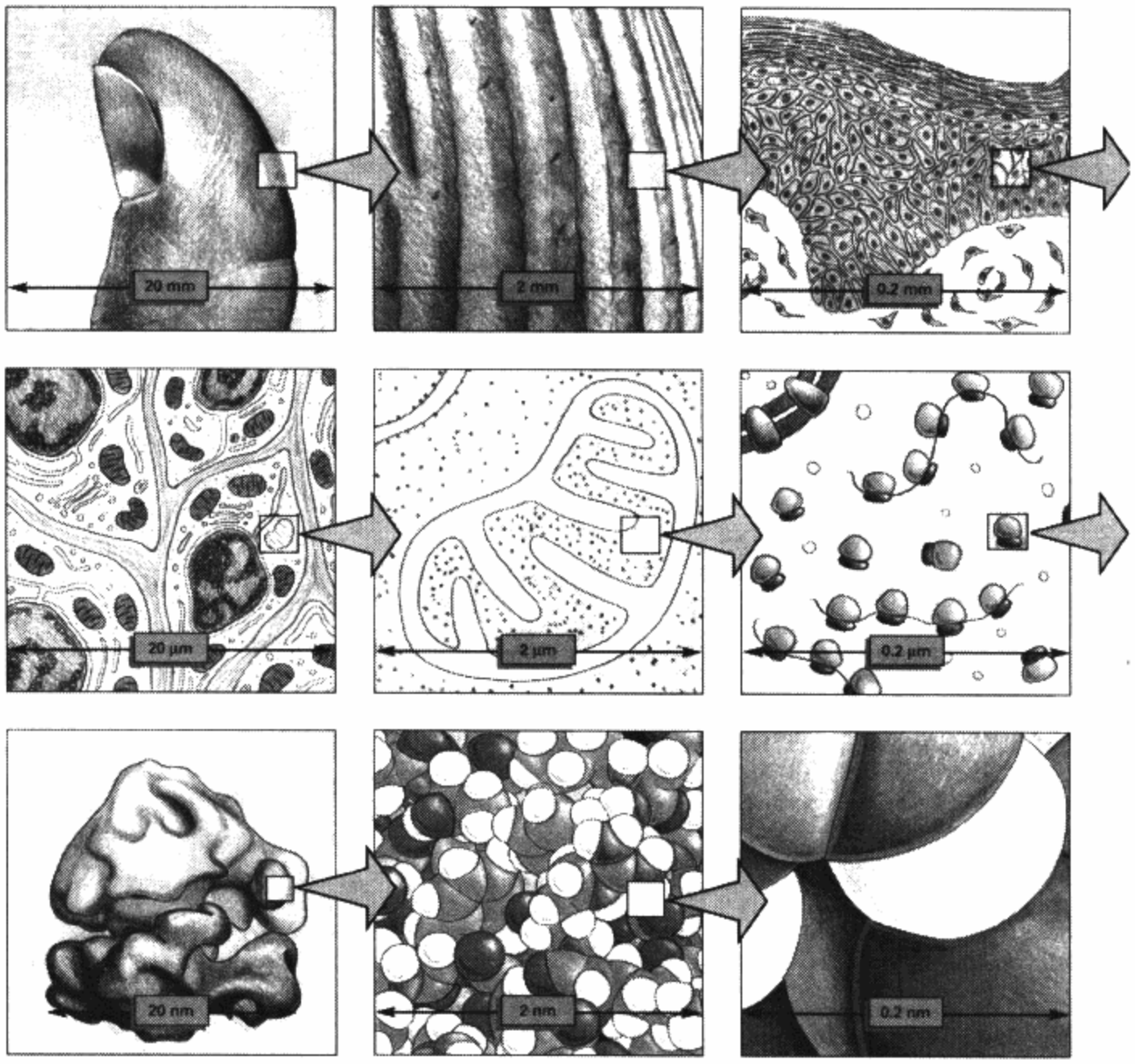


图 9-1　活细胞到原子之间大小尺度的直观图

将拇指到皮肤细胞再到核糖体再到作为体内蛋白质组成部分的一簇原子，用一系列虚拟的图像，按照每幅放大 10 倍的级数展示出来。显示在最后两格图像中的分子结构的细节，已经超出了电子显微镜的放大倍率

表 9-1　光学显微镜发展史中的一些重要发现

1611 年	Kepler 提出制作复杂显微镜的方法
1655 年	Hooke 使用复杂的显微镜，描绘了软木塞切片中的小孔，并将其称作细胞
1674 年	Leeuwenhoek 报道了他关于原生动物的发现。9 年后他首次看到了细菌
1833 年	Brown 发表了他对兰花的显微镜观察，明确地描绘出细胞核
1838 年	Schleiden 和 Schwann 提出细胞学说，宣称有核细胞是动植物结构和功能单位
1857 年	Kolliker 描绘了肌肉细胞中的线粒体
1876 年	Abbé 分析了显微镜中图像信息受衍射的影响，并指出显微镜设计如何最优化
1879 年	Flemming 描绘了动物有丝分裂中大而清晰的染色体
1881 年	Retzius 描绘了很多动物组织的细节，没有任何其他光学显微镜学家能超过他。在接下来的 20 年中，他和 Cajal 以及其他的组织学家发展了染色方法，奠定了微观解剖学的基础
1882 年	Koch 使用苯胺染料对微生物染色，鉴定出引起肺结核和霍乱的细菌。在接下来的 20 年中，其他微生物学家，如 Klebs 和 Pasteur，通过染色制备检测，鉴定了许多其他疾病的病原体
1886 年	Zeiss 参考 Abbé 的设计，制作了一系列的镜头，使得显微镜学家对结构的分辨达到了可见光的理论极限

续表

1898 年	Golgi 使用硝酸银对细胞染色，首次观察并描绘出高尔基体
1924 年	Lacassagne 和合作者首次发展出放射自显影法，使用放射性钋在生物标本中进行定位
1930 年	Lebedeff 设计并制造了第一台干涉显微镜。在 1932 年，Zernicke 发明了相差显微镜。这两大进展使得第一次观察到未染色的活细胞的细节
1941 年	Coons 使用偶联了荧光染料的抗体来检测细胞抗原
1952 年	Nomarski 设计微分干涉相差光学显微镜系统，并获得专利，现在这种显微镜仍以他的名字命名
1968 年	Petran 和合作者制造了世界首台共聚焦显微镜
1981 年	Allen 和 Inoué 完善了录像增强光学显微镜成像术
1984 年	Agard 和 Sedat 使用计算机去卷积重构了果蝇多线染色体核
1988 年	商业化共聚焦显微镜得到了广泛使用
1994 年	Chalfie 和合作者将绿色荧光蛋白（GFP）作为标记物引入显微镜技术中

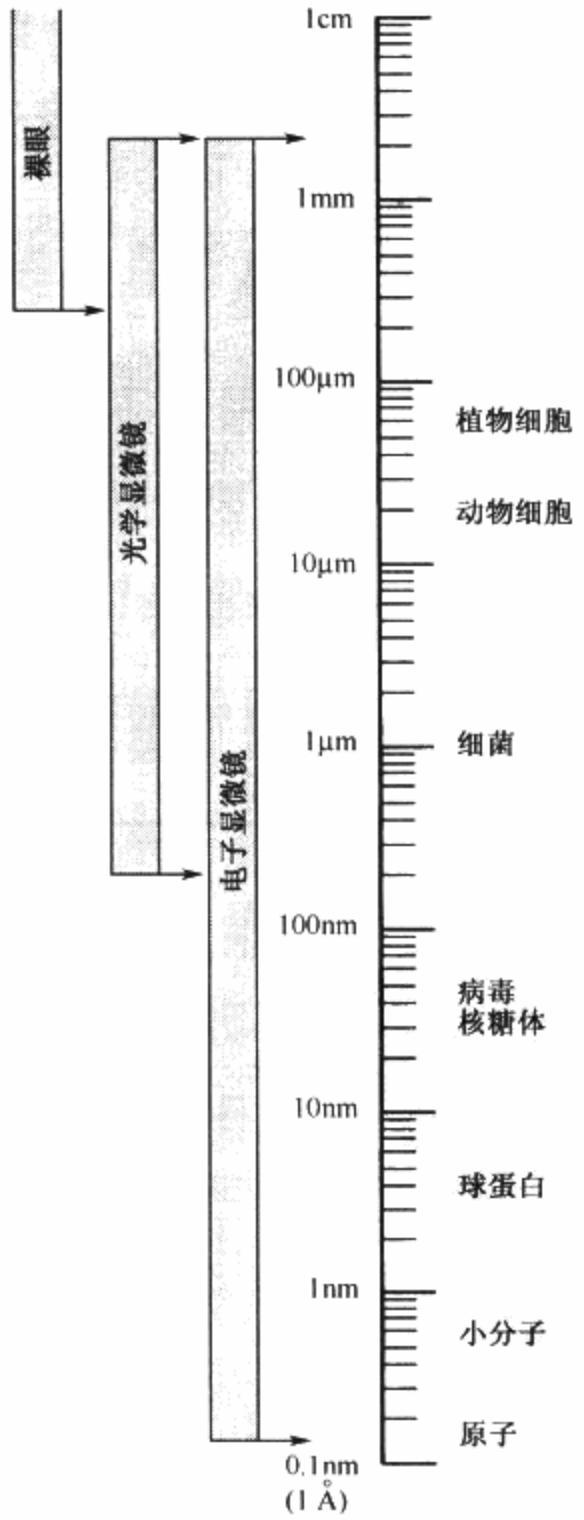


图 9-2　分辨能力

图中借助计算尺显示了细胞和其中成分的尺寸大小，也指出了通过肉眼、光学显微镜和电子显微镜分别能容易分辨的目标大小范围。下面是显微镜技术中常用的长度单位：

μm（微米）$=10^{-6}$m

nm（纳米）$=10^{-9}$m

Å（埃）$=10^{-10}$m

光学显微镜能分辨 0.2μm 的细节部分

通常，一特定的辐射不能被用于探测小于它自身波长的细微结构。这是所有显微镜的基本极限。因此，一个光学显微镜的最终检测极限取决于可见光的波长，这范围大概是从 0.4μm（紫色）到 0.7μm（深红色）。在实践中，宽约 500nm（0.5μm）的细菌和线粒体，通常是**光学显微镜**下能被清楚观测到的最小对象；由于光的波动性，比这更小的细节只能得到模糊的效果。要了解这种现象发生的原因，我们必须了解一束光通过显微镜的透镜时发生的现象（图 9-3）。

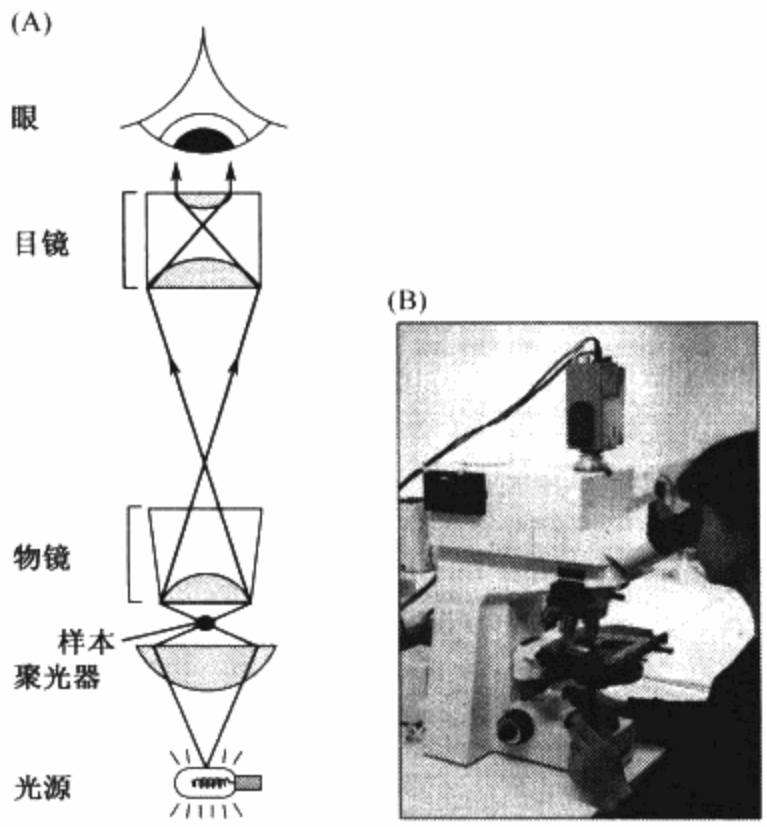


图 9-3　一台光学显微镜

（A）显示了一台复杂的显微镜中光通路，光线通过一聚光器中的透镜聚焦在标本上。物镜和目镜的组合通过调焦使得受照标本呈肉眼可见的像。（B）一台新式研究用光学显微镜

由于光的波动性，光并不像几何光学所预计的那样，严格按照理想的直线传播。光波反而按照轻微不同的多种路线穿过一个光学系统，以至于它们相互干涉，导致了光的衍射现象。若两列光经不同路径到达同一点且相位相同，波峰与波峰匹配，波谷与波谷匹配，这两列光将彼此增强变得更亮。相反，如果这两列光异相，它们将相互干涉，以至部分或完全地被抵消（图 9-4）。物体对光的干扰会改变光波的相位关系，产生复杂的干涉现象。例如，在高放大倍率时，一单波长的光照射一个物体直边缘得到的阴影是一列平行线，而照射一个圆点显示的阴影是一系列同心圆（图 9-5）。基于相同原因，通过显微镜看一单个的点，显示的是一个模糊的圆盘，而两个足够近的点显示重叠的图像甚至可能并作一个点。光学镜头再精巧也不能克服由光波动性决定的局限。

两个物体能够被区别开的间隔极限叫做**分辨限度**，这依赖于光的波长和使用的透镜系统的数字孔径。后面这个量和显微镜光线入口的宽度有关，可以依照显微镜和观测目标之间的距离来调节；也就是说显微镜眼睛睁得越大，看得越清楚（图 9-6）。使用紫光（波长 0.4μm）在最佳条件下有一个 1.4 的数字孔径，在光学显微镜中理论上可以获得低于 0.2μm 的分辨限度。在 19 世纪末时显微镜制造者达到了这个分辨率，不过在同时代工厂生产的显微

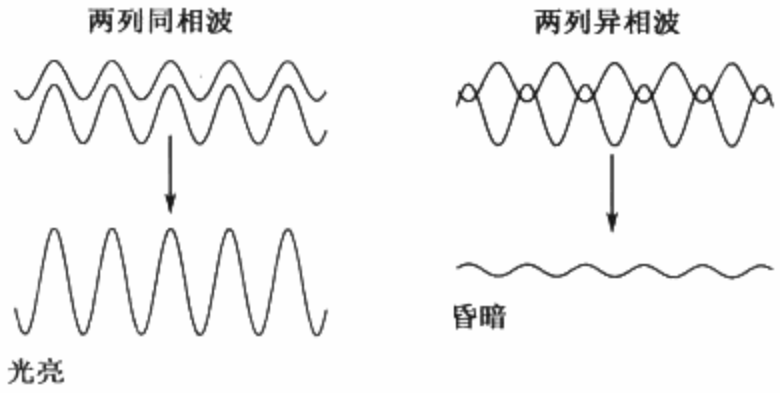


图 9-4　光波间的干涉

当两列同相光波合并时，合并后的波振幅增大，亮度增强。两列异相光波彼此抵消，产生的波振幅和亮度都减少

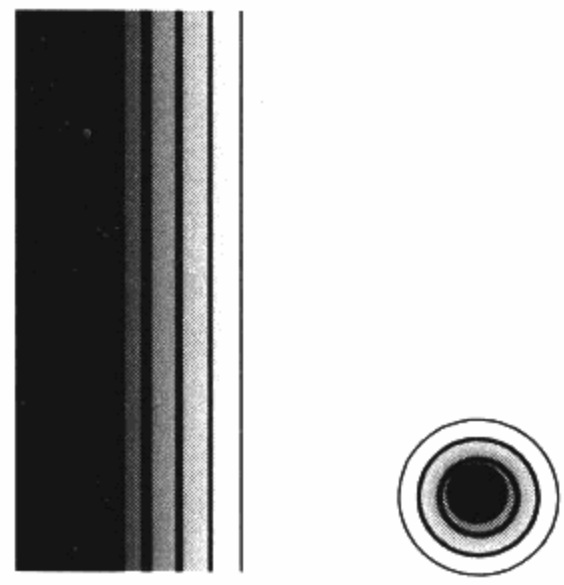


图 9-5　边缘效应

当光通过一放置于光源与观测器之间的致密物体边缘时，在高放大倍率下观测到的干涉效应

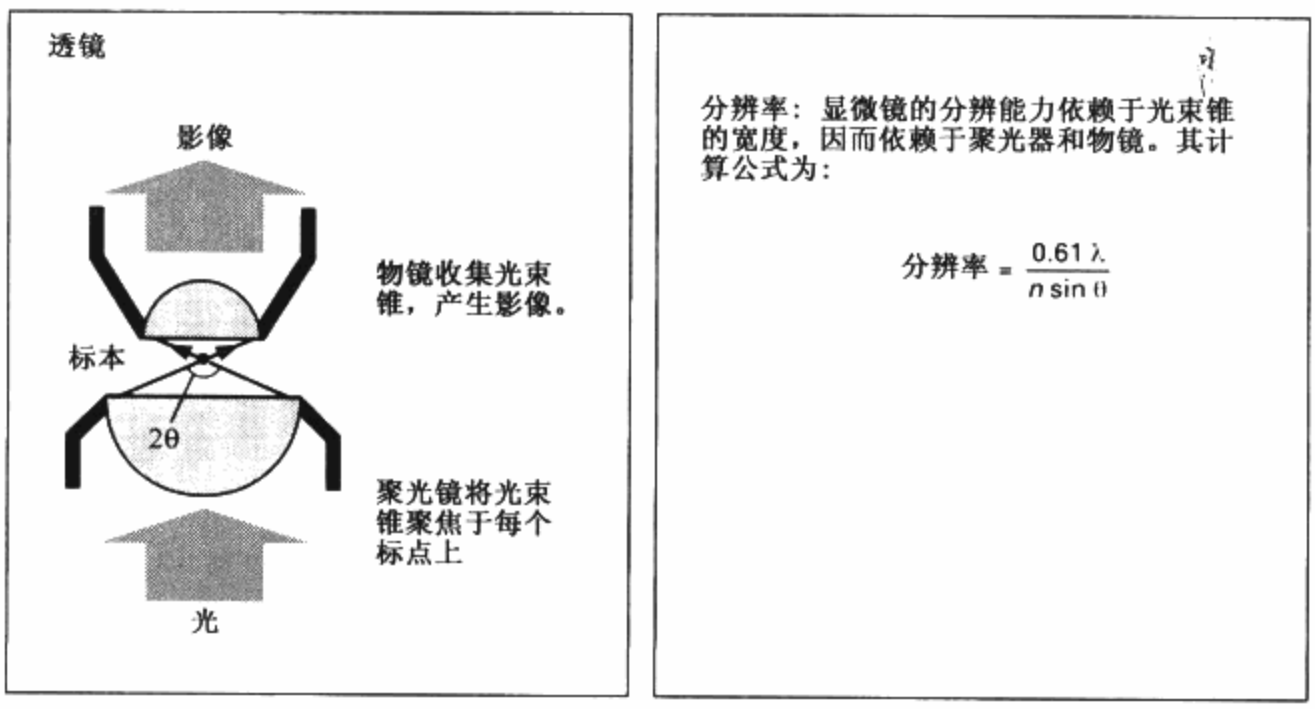


图 9-6　数值孔径

用显微镜中通过一透明标本的光线通路的图例来阐明数值孔径的概念，以及数值孔径与分辨率极限的关系

镜中很少能够达到这一分辨率。尽管像人们希望的那样放大图像是可能的，如将图像投影到屏幕上，但在光学显微镜中分辨两个间隔不到 0.2μm 的物体是不可能的；它们将显示为一个物体。

我们接下来看看干涉和衍射如何用于研究未染色的活细胞。然后我们详细了解如何固定细胞以用于光学显微镜的观察，以及细胞固定中的化学染料如何提高细胞结构的能见度。

用相差显微镜和微分干涉相差显微镜清晰地观察活细胞

在样品制备中细胞中某些组分可能丢失或者变形，这种可能性是对使用显微镜的技术人

员的挑战。当细胞是生活状态时进行观察，是唯一能避免这个问题的方法。为了这个目的，有着特殊光学系统的光学显微镜尤其有用。

当光穿过一个活细胞时，光波的相位根据细胞的折射率发生变化：光通过细胞中像细胞核这样相对较厚或较密的部分时被阻滞，因此它的相位相对于穿过周围较稀薄的细胞质的光，发生了改变。**相差显微镜**以及用一更复杂的方法的**微分干涉显微镜**，就是利用这样的两组光波重组时产生的干涉现象，来得到细胞结构的图像（图 9-7）。这两种光学显微镜都被广泛用于活细胞的成像。

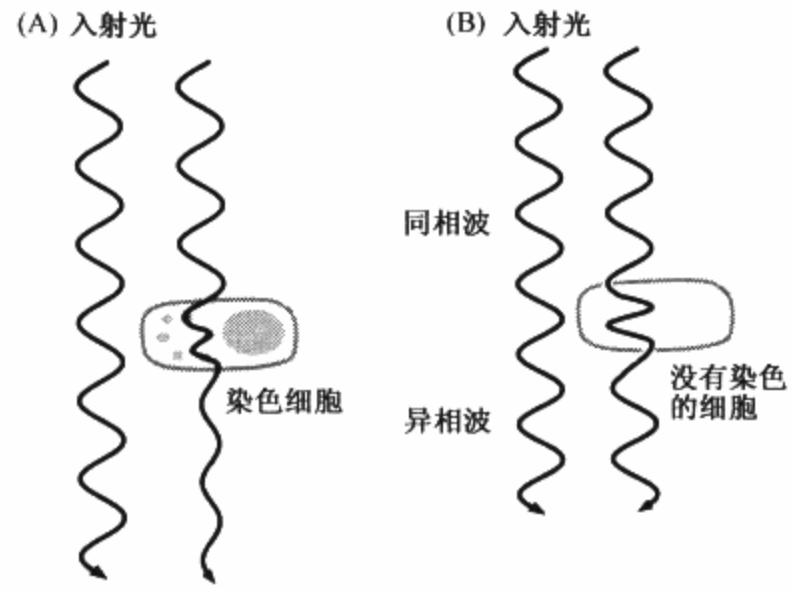


图 9-7　光学显微镜中两种获得相差的方法

（A）细胞中染色的部分会减少通过它们的特定波长的光波振幅，因此很明显可以按照常规方法获得细胞的彩色图像。（B）光通过未染色的活细胞在振幅方面经过很少变化，即使在高分辨率下也不能看到结构的细节。然而，在光通过细胞时，光的相位会发生变化，这种小的相位差别可以通过利用干涉效应借助相差显微镜或微分干涉相差显微镜变得可以看见

了解一个活细胞的某些特征的更简单方法是观察光线被细胞中多种成分散射的情况。暗视场显微镜中，入射光线从旁边射入，因此只有散射光才能进入显微镜镜头中。结果细胞作为明亮的物体显示在暗背景中。使用一个标准的**暗视场显微镜**，通过获得直接透过培养基中细胞的光线得到图像。图 9-8 显示了同一细胞在 4 种光学显微镜下获得的图像。

相差、微分干涉相差以及暗视场显微镜方法能让人观察到细胞有丝分裂以及细胞迁移等过程中的动作。由于许多细胞的动作太慢以至实时观察不到，因此常常采用拍摄间隔动作图像或者录像记录的方法。在这里，每隔一小段时间记录一帧图像，当按照正常速度播放记录的连续图像或者录像带时，事件显现的速度被大大加快了。

电子技术用于增强和分析图像

近年来电子成像系统和**图形处理**辅助技术对光学显微镜方法有重大影响。它们克服了显微镜应用中因为光学系统的缺陷而造成的某些应用限制。它们也避免了人眼的两个主要的弱点：在非常暗的光线下眼睛无法看清楚，以及在明亮背景下眼睛不能区分开细微不同的光线强度。第一个弱点可以通过给显微镜附加高感光摄影机（这类摄影机通常用于夜间监视）来克服。这使得在低光照水平下长时间观测细胞成为可能，从而避免了长时间亮光（和热）对

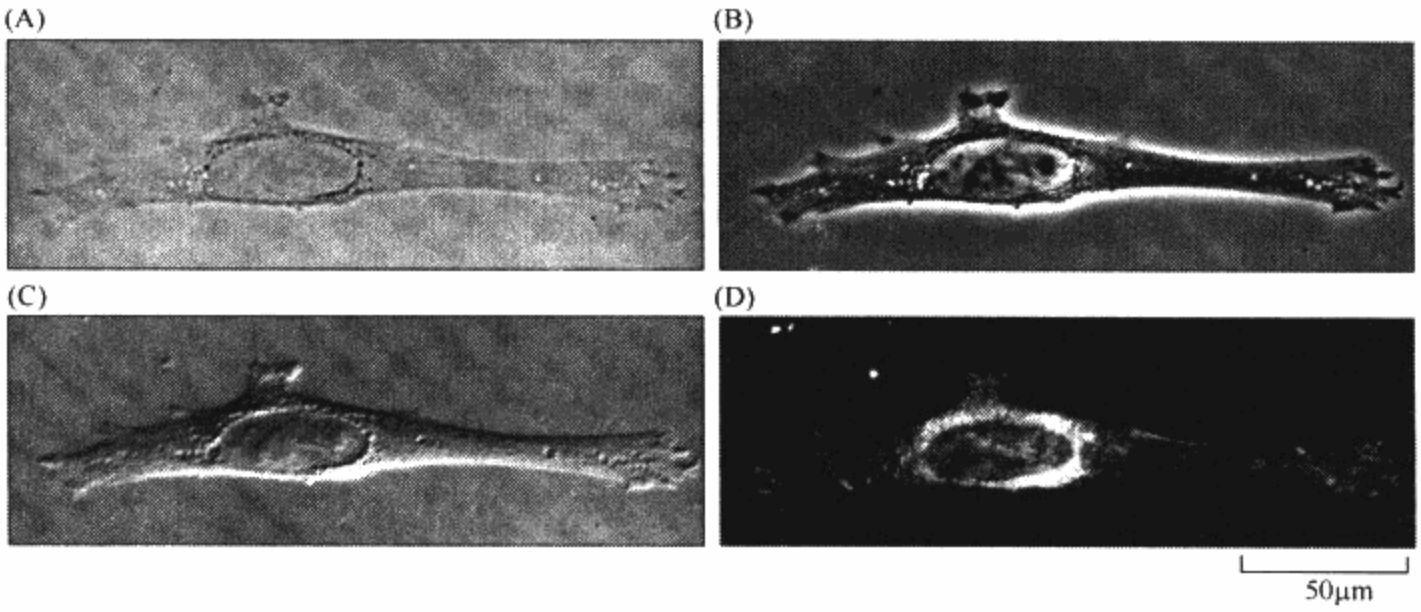


图 9-8　4 种类型的光学显微镜

4 幅图像分别显示了培养基中相同的纤维原细胞。大多数新式的显微镜通过交换光学元件都可以得到这 4 种类型的图像。（A）明视场显微成像；（B）相差显微成像；（C）Normarski 微分干涉相差成像；（D）暗视场显微成像

细胞的伤害。正如我们接下来要说明的，这种暗光线下的摄影机对于活细胞中的荧光分子的成像尤其重要。

由于摄影机获得的图像是电子形式的，这些图像很容易数字化，反馈给计算机，经过多种方法处理提取出隐藏的信息。这种图像处理过程有可能补偿显微镜中各种各样的光学缺陷，使得分辨率能达到理论的极限。此外，通过电子图像处理过程，相差能够被大大增强，这也克服了在辨别光亮度轻微差异方面肉眼的局限性。尽管这种处理过程也会增大光学系统背景不规则的效果，但是通过将视野中某个空白区域的图像作为本底，对整个图像进行扣除，能够克服这一缺点。那些小而透明的对象，原本不大可能从背景中区分开，经过上述处理后也变成可见的了。

计算机辅助的微分干涉相差显微镜具有高度反差，能够观察诸如单根微管（图 9-9）之类的非常微小的对象。单根微管直径 0.025μm，还不到光波波长的 1/10。在荧光显微镜下也能观察到荧光标记的单根微管（图 9-15）。不过在上述两种情况中，无法避免的衍射作用使得图像模糊，单根微管看上去至少有 0.2μm 宽，这也导致很难将单根微管从微管束中区分出来。

用于显微镜观察的组织通常要被固定和切片

要使细胞能被染色并能在显微镜下方便观察，第一步必须用固定剂处理细胞，固定住、杀死并保存好细胞。用化学术语来说，**固定**就是使得着色试剂能渗透入细胞并和细胞中大分子交联，使大分子变得稳定并固定在原位。早期的某些固定过程包括将样品浸入酸液或乙醇等有机溶液中。现在的过程通常包括用活性醛来处理，常用甲醛和戊二醛，这些醛能和蛋白质的自由胺基形成共价键，进而和周围的蛋白质分子交联。

大多数组织样品太厚了以至于不能在高分辨率下直接观察到单个的细胞。不过，固定后通常用切片机将组织切成非常薄的切片，切片机具有锋利的刀片，工作原理类似于普通的肉类切片机（图9-10）。然后将切片（通常 1～10μm 厚）平放在载玻片上。

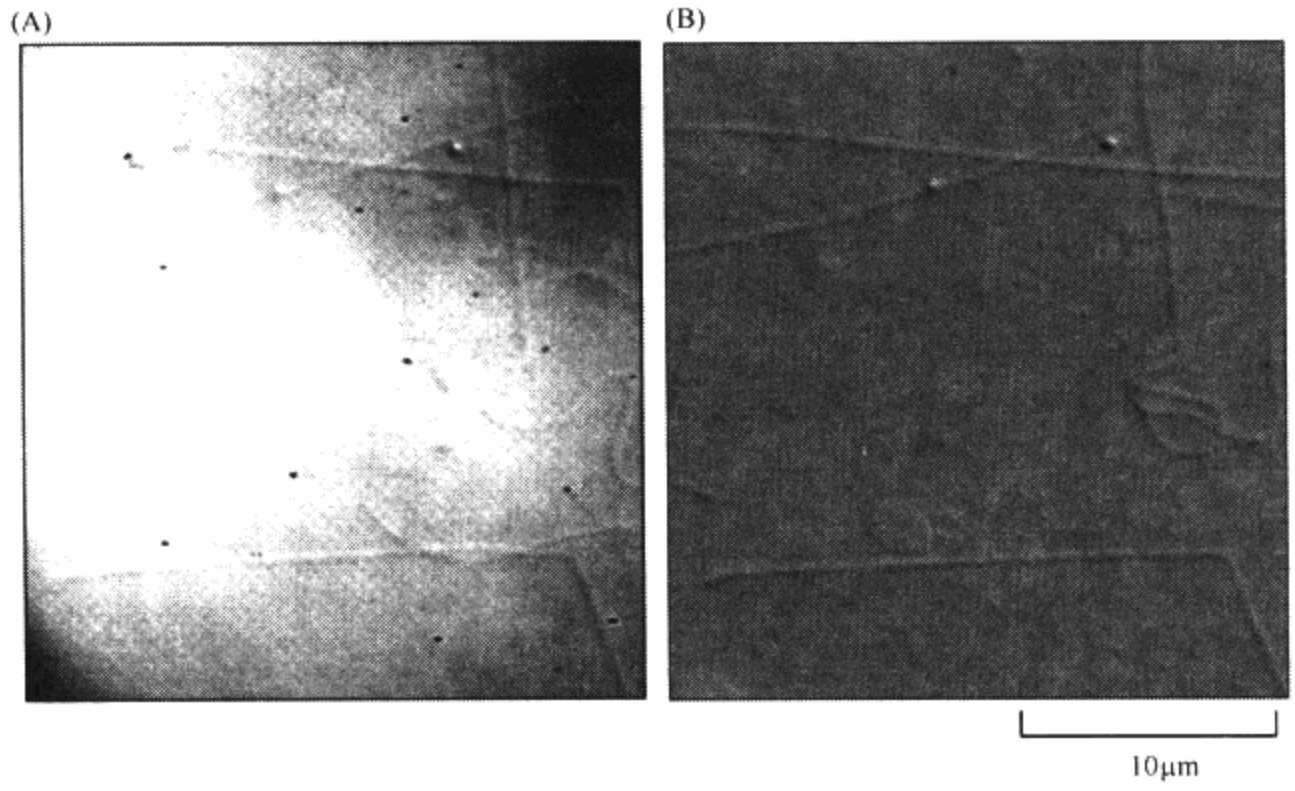


图 9-9　图像处理

(A) 显示的是一个未染色的微管没有被处理过的数码图像，由微分干涉相差显微技术获得的。(B) 图像现在经过处理，首先通过数位消减不均匀的亮背景，接着通过数位增强反差。图像处理的结果是得到一个更能说明问题的照片。需要注意的是微管是动态的，在前后图像之间有些已经改变了长度或位置（经 Viki Allan 许可）

即便在被固定后，组织仍通常是软的而且易碎，因此在切片前需要将组织包埋在支持物中。通常用石蜡或树脂作为包埋物。这两种材料在液态时都能围绕并渗入固定好的组织；这些材料凝固（冷却或聚合）后形成固块，很容易用切片机切片。

通常的固定或者包埋处理方法可能会有改变细胞的结构或者组分分子的风险。快速冰冻提供了另一种样品准备方法，通过避免固定和包埋，一定程度上避免了上述风险。通常在冰室中用特殊的切片机直接对冷冻组织进行切片。尽管用这种方法得到的冷冻切片避免了一些人为假象，但也遇到了问题：尽管像蛋白质之类的分子的天然结构得到了很好的保护，但细胞的细微结构常常会被冰晶破坏。

无论用哪种方法进行组织切片后，接下来的步骤通常是染色。

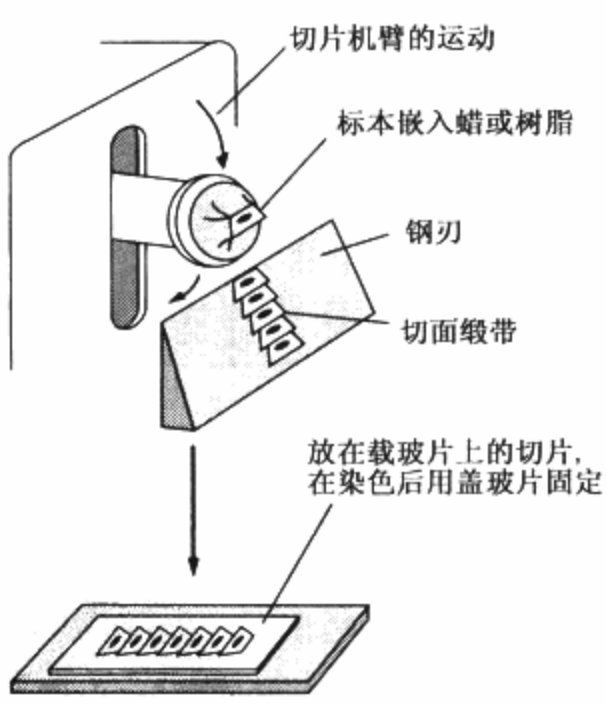


图 9-10　制备组织切片

插图显示了如何使用超微切片机将包埋好的组织进行切片以用于光学显微镜检测

细胞中不同的成分可以被选择性染色

大多数细胞的内容物（细胞全重的 70%是水）几乎不会阻挡光线的通过。因此，大多数细胞在自然状态时，即使是被染色或切片，在普通的光学显微镜下几乎也是不可见的。用染料进行染色是使细胞变得可见的一种方法。

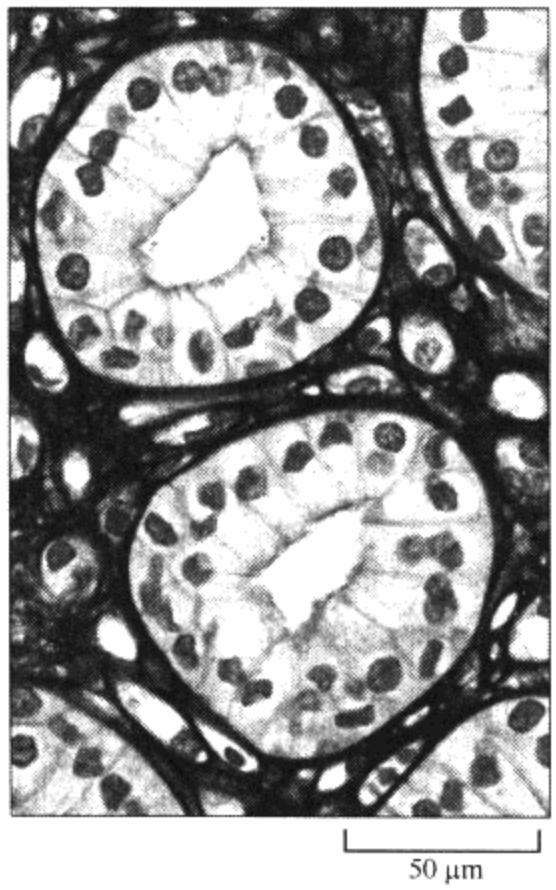


图 9-11　一个染色的组织切片

肾的集尿管中细胞的切片被用苏木精和曙红这两种组织学中常用染料的混合物染色。每个管是由相互紧密排列的细胞（细胞核被染成红色）组成，这些细胞形成一个环。这个环被胞外基质（染成紫色）包围（摘自 P. R. Wheater et al., Functional Histology, 2nd edn. London: Churchill Livingstone, 1987）

在 19 世纪早期，对印染纺织品的染料需求导致了有机化学的繁荣时期。某些染料被发现能用于生物组织染色，而且出乎意料地在细胞中的某些特殊部分，如细胞核和线粒体中被优先显示出来，这使得细胞内部结构清晰可见。现在被使用的有机染料有很多，根据颜色命名的有孔雀绿、苏丹黑以及考马斯蓝等，每一种染料对特殊的亚细胞组分有着特殊的亲和力。例如，苏木精染料对带负电的分子有亲和力，因此可以用于显示 DNA、RNA 以及酸性蛋白质在细胞内的分布（图 9-11）。不过对于许多染料特异性的化学基础仍然不清楚。

对于这些染料在分子水平的特性了解的相对缺乏，促使人们寻找更多合理的染色方案、选择性的染色过程和特殊方法来显示细胞中某一特定的蛋白质或大分子。但为了达到这种效果仍然存在一个问题。因为在任何特定的细胞中，大多数大分子的个数往往不多，每个大分子上结合一到两个染料分子后通常并不能被观察到。解决这一问题的方法是增加单个大分子上结合染料分子的个数。不过，有些酶可以根据它们自身的催化活性在细胞中定位：当使用合适的底物分子时，每个酶分子能催化很多底物分子反应，得到局部分布的可见的反应产物。另外，我们也可以选择使用荧光染料来解决灵敏性的问题，接下来我们会详细介绍这种方法。

通过荧光显微技术对特定分子进行细胞定位

荧光分子吸收某一波长的光然后发射出另一波长更长的光。若用某化合物吸收波长的光照射它，然后通过一个只允许发射波长的光通过的滤光片来观察，这样看上去就是在一黑暗背景中发光。由于背景暗，即使很微量的荧光染料发出的光也能被探测到。而相同数量的普通染料分子只能使得光在通过样品被染色的部分时轻微的改变颜色，通常是无法被观测到的。

用于细胞染色的荧光染料可直接用**荧光显微镜**进行观测。这种显微镜和普通的光学显微镜比较类似，不同之处在于从强光源发出的光需要经过两套滤光片，一套用于过滤到达样品之前的光，

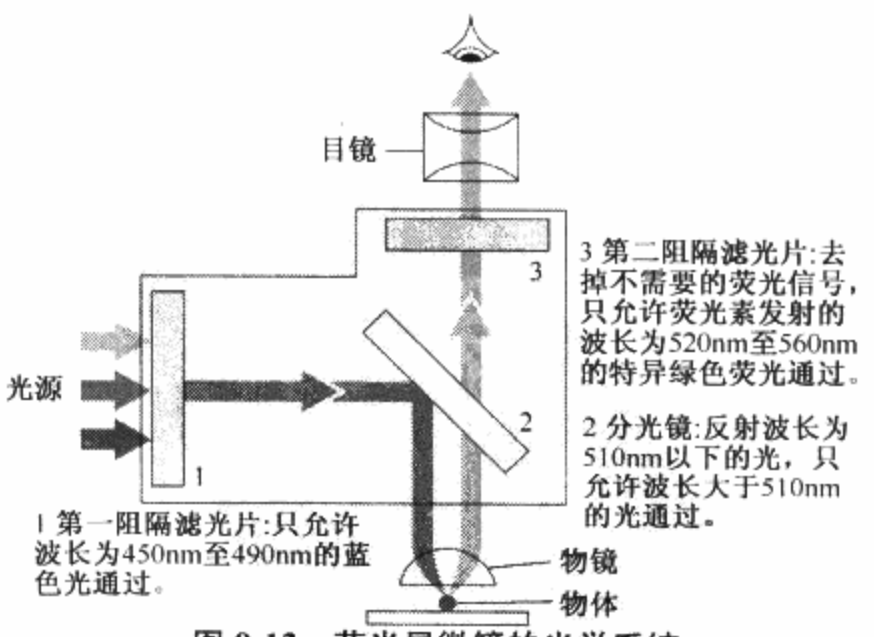


图 9-12　荧光显微镜的光学系统

一套滤光装置由两个光栅（1 和 3）和一个将光束一分为二的分色镜（2）组成。这里显示的是用于检测发荧光的荧光分子的滤光装置。在这类显微镜方法中高数值孔径的物镜尤其重要，对于某一特定放大倍率，荧光图像的亮度和数值孔径的 4 次方呈正比（图 9-6）

另一套用于过滤经过样品之后获得的光。第一套滤光片选择性地获得荧光染料激发波长的光，而第二套滤光片就完全屏蔽掉这些光线，只让荧光染料发射波长的光通过（图 9-12）。

荧光显微方法主要被用于检测细胞和组织中的特异蛋白质或者某些分子。一种已被广泛使用的有效方法是将荧光染料偶联到抗体分子上，这些抗体作为高特异性的、通用型着色剂，能够选择性地和细胞中或细胞外基质中它们所识别的特定的生物大分子结合。荧光素和罗丹明是这类荧光染料中最常用的两种，荧光素被蓝光激发后发出强烈的绿光，罗丹明被黄绿色的光激发后发出深红色的荧光（图 9-13）。将荧光素偶联到一种抗体上，罗丹明偶联到另一种抗体，可以用来比较细胞中不同分子的分布情况；通过切换前后两套分别针对不同染料的滤光片，这两种分子可以在显微镜下分别显现。正如图 9-14 显示的，用同样的方法，可以使用 3 种荧光染料来区别开同一细胞中的 3 种类型的分子。许多新的荧光染料，如 Cy3、Cy5 以及 Alexa 染料，也已经用于荧光显微方法中（图 9-13）。

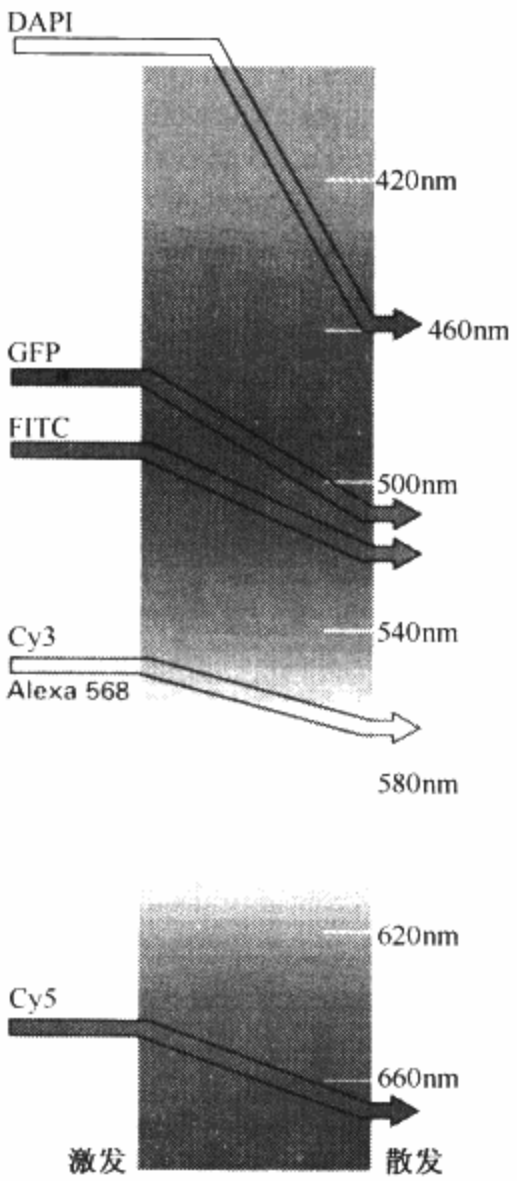


图 9-13　荧光染料

图中显示了几种常用的荧光染料的最大激发和发射的波长，以及这些波长对应的光的颜色。染料分子放出的光子的能量必须要低于所吸收光子的能量，也要低于激发和发射峰之间能量的差值。GFP 是一种绿色荧光蛋白，不是一种染料。DAPI 是一种广泛使用的常规荧光 DNA 染料，吸收紫外光发出明亮的蓝色荧光。其他的染料通常用于抗体和其他蛋白质的荧光标记

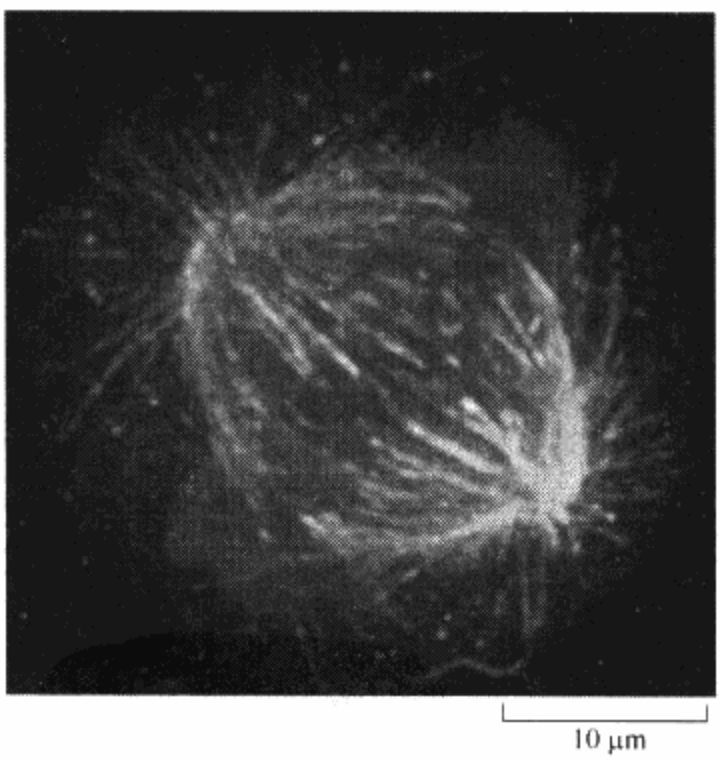


图 9-14　多荧光探针显微技术

处于有丝分裂中的细胞成分显微照片，用 3 种不同荧光探针显示 3 种不同的细胞成分。纺锤体微管、着丝粒和浓缩染色体的 DNA 分别用绿色的荧光抗体、红色的荧光抗体和蓝色的荧光染料 DAPI 显示（经 Kevin F. Sullivan 许可）

在本章稍后部分论述的重要方法中，还要介绍荧光显微技术如何用于检测活细胞中某些分子浓度和位置变化。

抗体可用于检测特定分子

抗体是脊椎动物免疫系统抵御感染时产生的蛋白质（讨论见第 24 章）。它们在蛋白质中是很独特的，因为抗体有数以 10 亿计的不同形式，每一种有一不同的结合位点能够识别某一特异的靶标分子（或抗原）。抗体这种严格的抗原特异性使得它们成为细胞生物学家强有力的工具。用荧光染料标记后的抗体，常常被用于荧光显微方法中特定分子的细胞内定位（图9-15）；而对于那些使用胶体金之类的电子密度高的颗粒来标记的抗体，通常在电镜下有类似的应用。

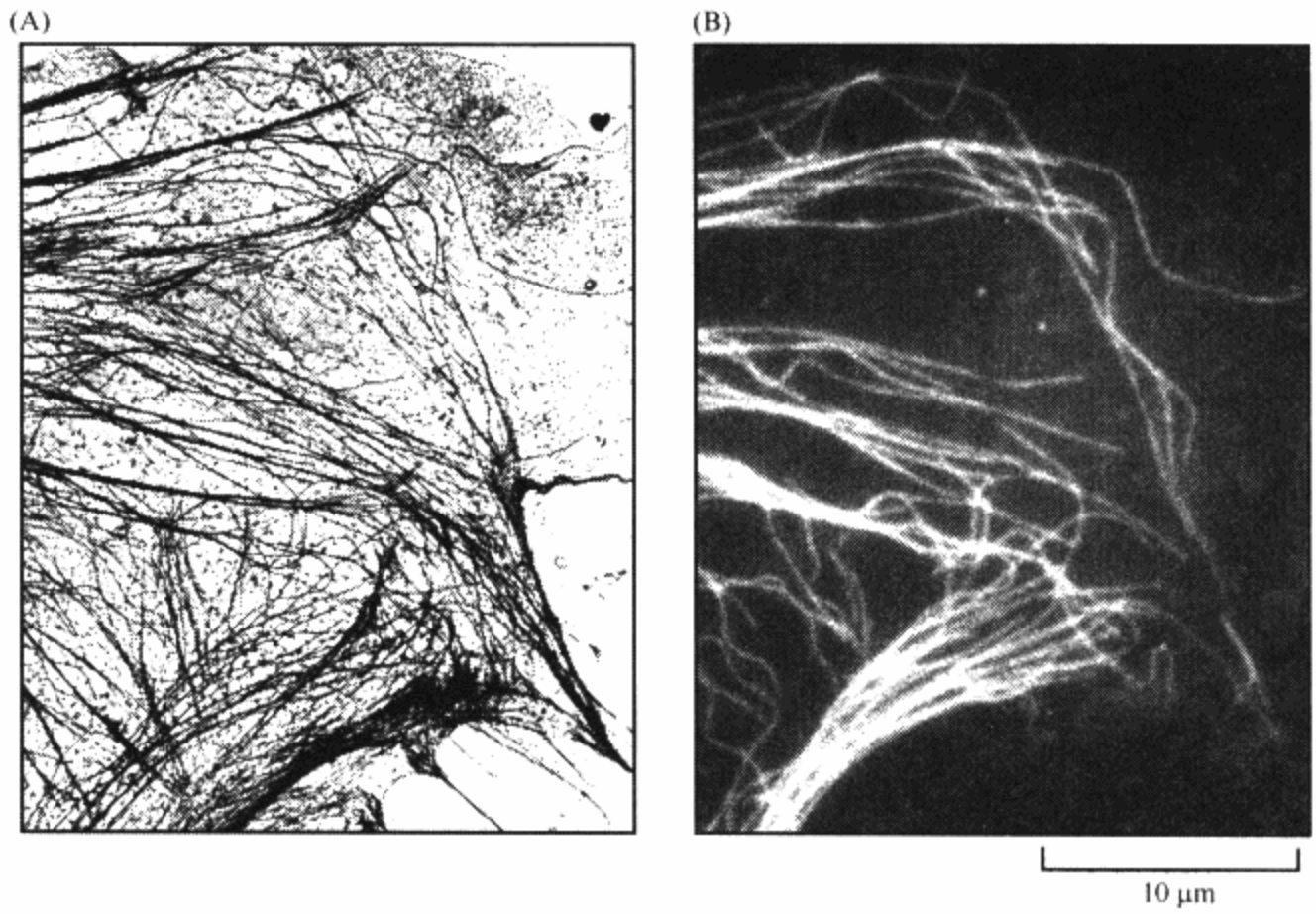


图 9-15　免疫荧光技术

（A）一个培养的上皮细胞的透射显微照片显示了微管和其他纤维的分布。（B）用抗微管蛋白（微管的蛋白质亚基）的荧光抗体这一间接的免疫细胞化学技术染色同样区域（图 9-16）。箭头指示的是容易识别的单个微管。需要注意的是因为衍射效应，光学显微镜下微管显示为 0.2μm 宽，比它们实际的 0.025μm 的宽度要宽（图片摘自 M. Osborn, R. Webster, and K. Weber, *J. Cell Biol.* 77: R27～R34, 1978. © The Rockefeller University Press）

通过一些化学方法来增强信号能够提高抗体在检测和分析细胞以及组织中的特定分子时的灵敏度。例如，尽管能够将荧光素等标记分子直接偶联到抗体上作为特异性识别的一抗来使用，但不如先用未标记的一抗，然后用标记好的二抗来和一抗结合，这样能得到更强的信号（图 9-16）。

最敏感的信号放大方法是使用酶作为标记分子偶联到二抗上。例如，碱性磷酸酶能够在适当的化学物存在时，产生无机磷酸盐，在局部形成有色沉淀。这能够显示出酶联二抗的位置，因此也能显示出二抗结合的抗原抗体复合物的位置。由于每个酶分子能够催化产生成千上万的产物分子，因此即使是很微量的抗原也能被检测到。基于这种原理，酶联免疫分析（ELISA）作为灵敏的检测方法常被用于医学检测中，如用于检测怀孕以及多种传染病。尽管酶的放大效应使得酶联方法灵敏度非常高，但是有色沉淀的扩散意味着这种方法在显微镜下空间定位的局限性，荧光标记常被用于精确的空间定位。

抗体制备很简单，将抗原样品注射到小鼠或山羊等动物中几次之后，收集这些动物富含抗体的血清即可。抗血清中包含不同种类的抗体的混合物，每种抗体都由一不同的抗体分泌

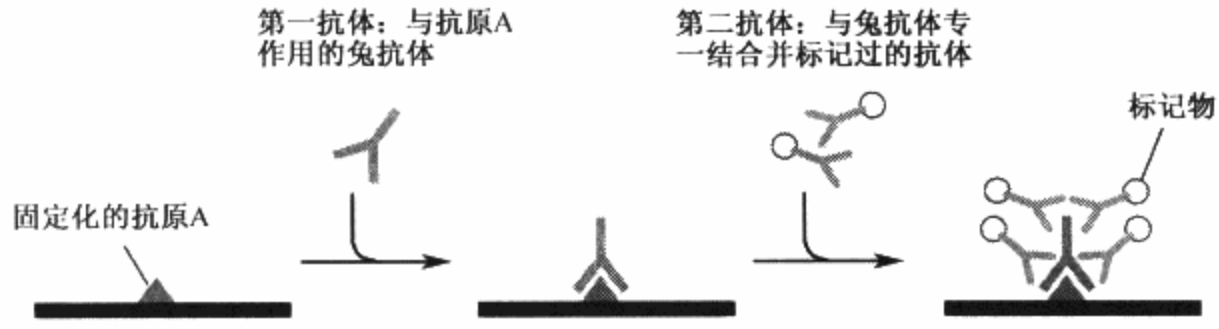


图 9-16　间接免疫细胞化学

这种检测方法非常敏感，因为一抗本身能被许多二抗分子识别。二抗上共价连接一个标记分子，使之易于被探测到。常用的标记分子包括荧光染料（用于荧光显微成像）、辣根过氧化物酶（用于常规的光学显微成像或电子显微成像）、胶体金颗粒（用于电子显微成像）以及碱性磷酸酶或过氧化物酶（用于生化测定）

细胞（B 淋巴细胞）产生。不同的抗体识别抗原不同的部分（这叫做抗原决定簇或表位）或抗原制剂中的杂质。要提高某一特定抗原的抗血清的特异性，有时可以靠除去其中那些结合其他分子的抗体来达到。例如，某一抗蛋白质 X 的抗血清，可通过一个连有蛋白质 Y 和蛋白质 Z 的亲和柱来除去那些抗 Y 和抗 Z 的抗体。即便如此，抗血清中的不同成分有时还是限制了它们的有效性。单克隆抗体的使用很大程度上解决了这个问题（图 8-6）。不过单克隆抗体也有些问题。因为单克隆抗体是单一种类的抗体蛋白，它们对抗原的单一位点或决定簇有很好的特异性，但是抗体有效性也由抗原表位的易接近性决定，这取决于样品的制备过程。例如，有些单克隆抗体只和未固定的抗原作用，另一些只和经过特殊固定后的抗原作用，还有的只和在 SDS-PAGE 中变性的蛋白质作用，并不能和蛋白质的天然结构发生作用。

使用光学显微镜进行复杂的三维物体成像是可能的

正如我们所知道的，使用普通的光学显微镜来观察组织，需要将组织切成薄片供检查；切片越薄，图像越明晰。在切片的制备中，第三维的信息会丢失。那么怎样才能获得细胞或组织的三维结构图像，对于那些由于某种原因不能切片制片的样品怎样观察它们的显微结构？尽管光学显微镜在观察复杂的三维标本时能够聚焦在一个精确的焦平面上，但这个标本焦平面以上或以下的部分也会被照亮，这些区域散射出的光使得图像出现模糊的光斑。这种焦点之外的光使得要阐明图像的细节变得很困难，也会模糊好的图像构成。

已经发展出两种方法来解决这个问题：一种是计算的方法，另一种是光学方法。三维显微成像方法通过滤去焦平面外的光，能够在厚标本中聚焦到某选定平面。因此可以看到明晰的薄的光学截面。提取不同深度的光学截面存储到电脑中，很容易重构出三维图像。这种方法对于那些用 CT 扫描仪检测人体的研究人员非常有用：这些机器都能给出内部完整结构的详细图像。

这种计算的方法通常被叫做图像去卷积。要了解它如何运作，需要回忆光的波动性是如何使得点光源通过显微镜透镜系统成像为一个小的模糊斑，如果点光源位于焦平面的上方或下方会增大图像的不清晰。点光源的这种模糊图像被称作点扩展函数。对于一个完全模糊的图像，可以用相应的模糊斑来当作标本中的每个点，然后建立一个复杂物体的图像。对于去卷积（deconvolution），我们首先通过显微镜聚焦在一系列焦平面上，获得一系列（模糊的）图像，得到模糊的三维图像效果。然后通过计算机处理这些图片，尽可能的减少模糊。这计算机程序本质上是使用显微镜点扩展函数来测定图片上的模糊效果，然后应用等效的去模糊化（去卷积化），把模糊的三维图像转化为一系列清晰的光学截面。这需要的计算十分复杂，

而且使用时有一系列限制。不过随着计算机越来越快也越来越廉价，图像去卷积的方法正变得更有力也更普遍。具体例子如图 9-17 所示。

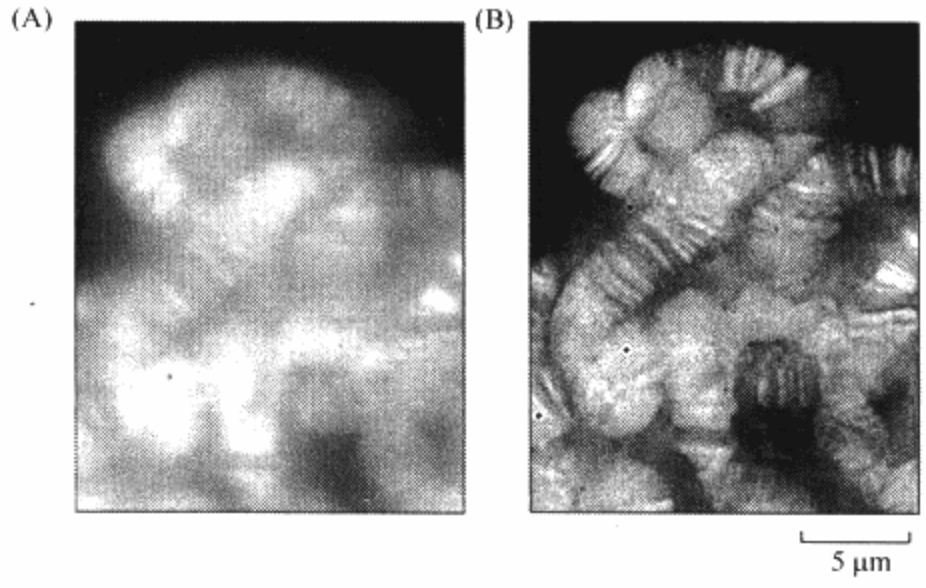


图 9-17　图像去卷积

（A）果蝇的多线染色体的光学显微照片，用荧光 DNA 结合染料染色。（B）图像去卷积后同一区域的图像清楚地显示了染料在染色体上的结合图案。每条带大约 0.25μm 厚，接近了光学显微镜分辨率极限（经 John Sedat Laboratory 许可）

共聚焦显微镜通过滤去焦平面外的光获取光学截面

共聚焦显微镜能达到去卷积化相似的结果，不过在测量前需要对光进行处理；这是一种比数字化更好的技术。**共聚焦显微镜**的光学元件很复杂，但基本原理是简单的，如图 9-18 所示。

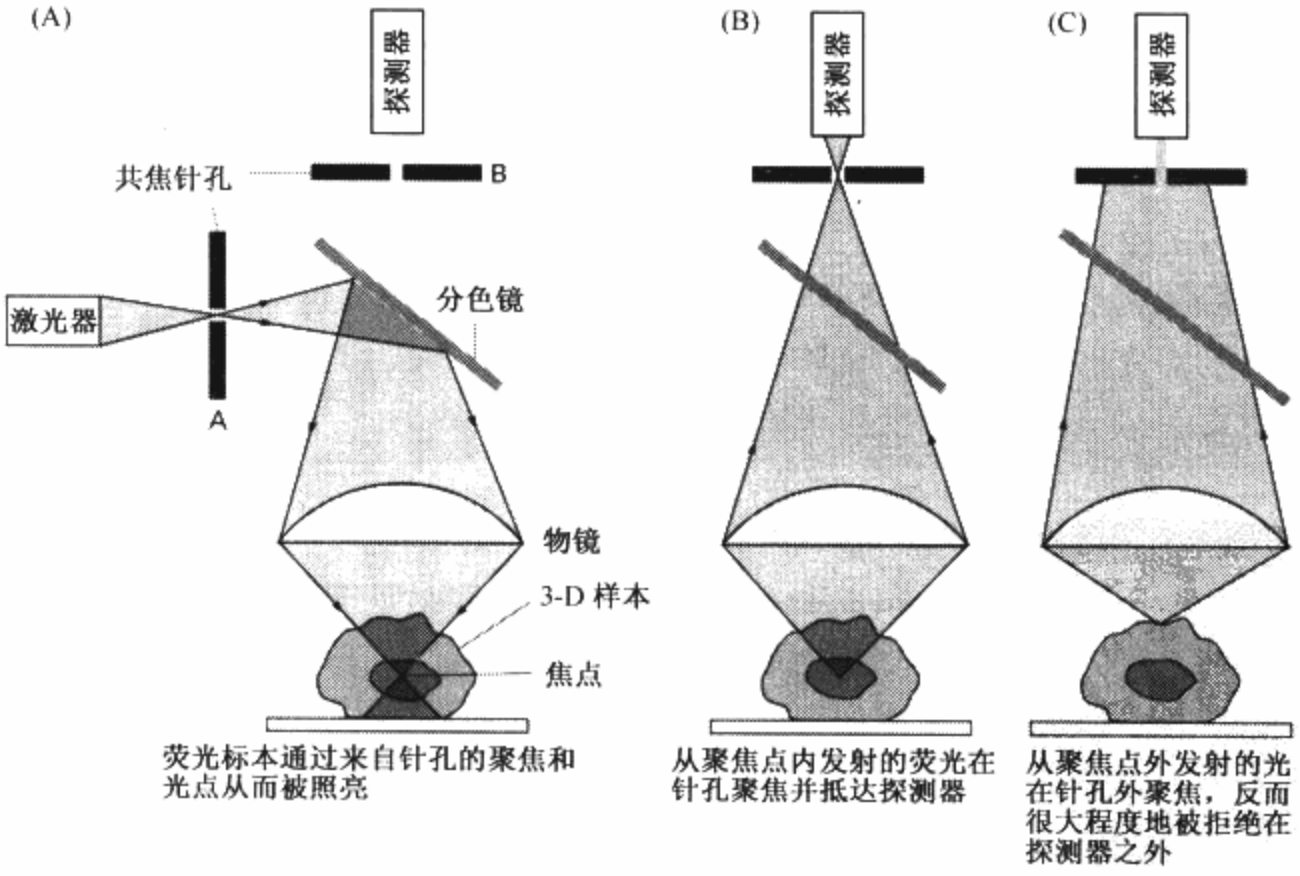


图 9-18　共聚焦荧光显微镜

简图显示这种显微镜的光学元件基本排列和图 9-12 中显示的标准的荧光显微镜相似，除了用一束激光照亮一个小针孔，针孔的图像能在样品中聚焦成一个点（A）。从样品中的焦点发出的荧光在第二个针孔处被聚焦（B）。从样品其他部分发出的光不能聚焦在这里，因此不会出现在最终的图像中（C）。用穿过样品的光束进行扫描，可以获得样品中焦平面的清晰图像，这个图像不会受到来自样品其他的区域的光干扰

这种显微镜通常使用荧光系统（图 9-12），但不是一次照亮整个标本，用通常的方法，光学系统瞬时将少量光聚焦到样品某个深度的一个点上。这需要一个非常亮的微小光源；这通常使用通过针孔的激光来提供。从被照亮的材料发出的荧光可以被收集和利用适当的光学探测器来得到图像。针孔开口被安置在探测器前，这个位置和入射光针孔是共焦的，也就是说从样品中被照亮的点发出的光线在那个位置能够聚焦。因此在样品中来自这个点的光在针孔开口聚焦后进入探测器。

相对应的，来自于聚光灯焦平面之外区域的光在针孔开口处焦点也是没对准的，因此大部分进不到探测器（图 9-19）。要逐步建立一个二维图像，可通过使用光栅扫描这个焦平面区域顺序得到每个点的数据，收集这些数据，显示在电视屏幕上。虽然没有显示在图 9-18 中，这种扫描通常通过摆动置于分色镜和物镜之间的镜子来偏转光束，入射聚光灯和探测器的共聚焦针孔按照这样一种方式保持严格的记录。

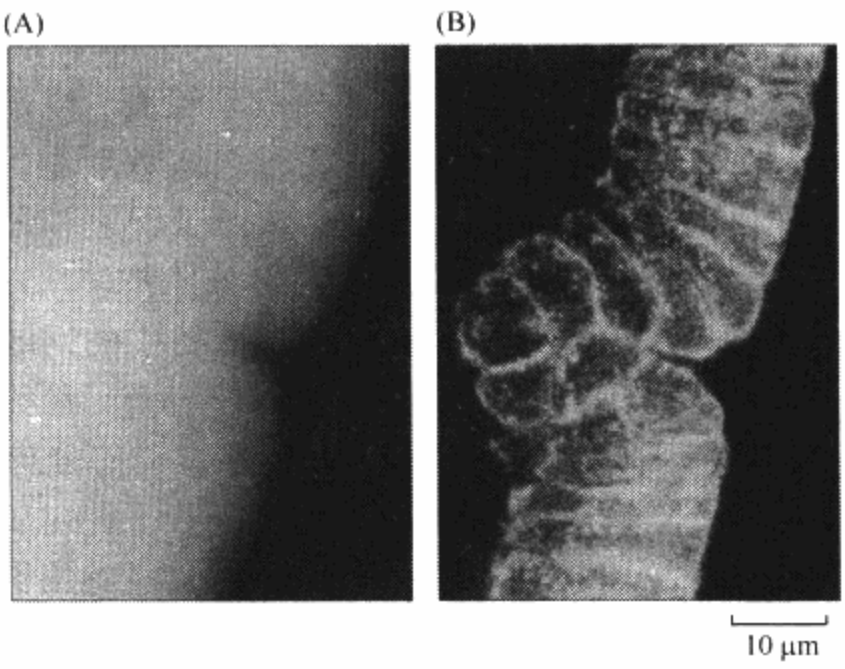


图 9-19　常规荧光显微镜技术和共聚焦荧光显微镜技术的比较

这两张显微照片是关于同一个囊胚期果蝇胚胎的，使用了针对肌动蛋白纤维的荧光探针来染色。（A）常规的、未处理过的图像被焦平面上下存在的荧光结构模糊。（B）在共聚焦图片中，焦平面之外的信息被除掉，获得胚胎细胞的一个轮廓清晰的光学截面（经 Richard Warn 和 Peter Shaw 许可）

共聚焦显微镜已经被用于了解许多复杂三维对象的结构（图 9-20），包括细胞质中细胞骨架纤维网络和细胞核中染色体和基因的排列。

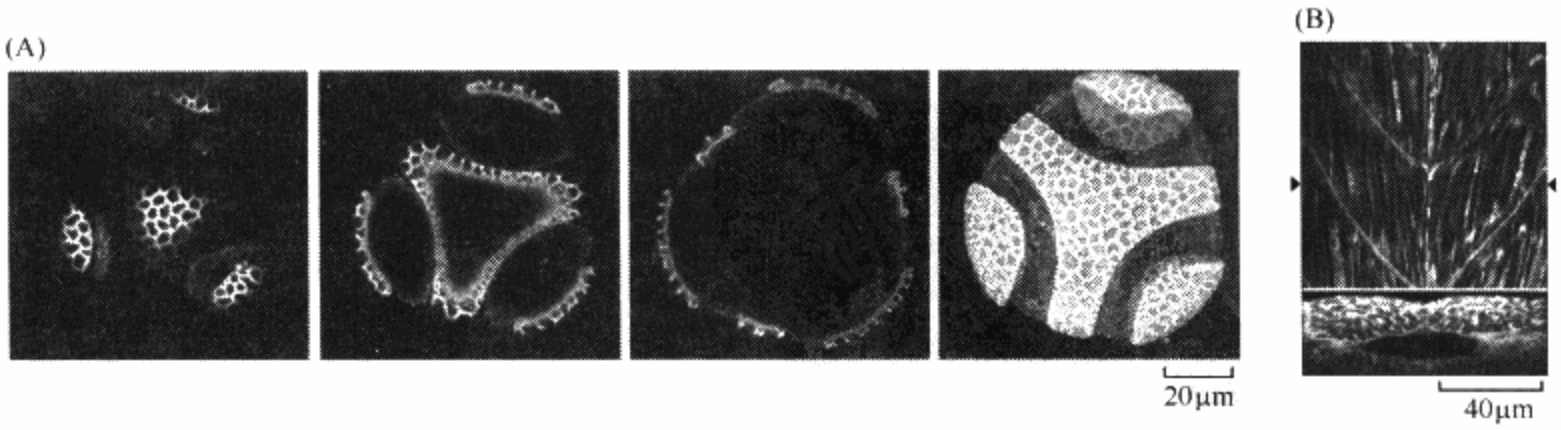


图 9-20　用激光共聚焦显微照片进行三维重构

（A）西番莲的花粉粒具有复杂的巨刻纹的细胞壁，其中包含荧光成分。使用共聚焦显微镜从花粉粒不同深度获得的图像能够被重组成一个完整花粉的三维图像。从全套 30 个截面（这些界面和它的邻居间有点重叠）中选出的 3 个单独的光学截面显示在左边。（B）斑马鱼尾部区域用两种荧光染料染色，在共聚焦显微镜下成像。下面的图像是尾巴的一个光学横截面，这一截面是当焦平面的逐渐提高时，激光在单线上扫描获得的数位构造。中间的暗腔是正在发育的脊索［（A）经 Brad Amos 许可；（B）经 S. Reichelt 和 W. B. Amos 许可］

去卷积法和共聚焦显微成像作为三维光学显微成像术的优缺点仍然是辩论的主题。共聚焦显微镜通常比去卷积系统更容易使用，能够快速观察到最终的光学截面。另一方面，用于去卷积化系统的新型冷却式 CCD（电荷耦合设备）相机对于收集少量的光非常有效，它们能用于制作那些不容易染色或容易被共聚焦技术中强光伤害的标本的详细的三维图像。

电子显微镜分辨细胞的细微结构

对于任何形式的射线来说分辨限度和照射波长之间具有严格的相关性（图 9-6），无论是一束光还是一束电子。不过使用电子时分辨限度可能达到非常小。一个电子的波长随着电压增加而减小。在一个有着 10 万伏加速电压的**电子显微镜**中，电子波长是 0.004nm。这样一台显微镜理论上的分辨率应该大约为 0.002nm，这是光学显微镜的10 000倍。相比那些玻璃镜头，电子显微镜镜头的偏差更难被校正，不过大多数新型的电子显微镜的实际分辨率能达到 0.1nm（1Å）（图 9-21）。因为电子镜头只有正中间的部分才被使用，有效的数字孔径非常小。此外，标本准备中的问题和辐射伤害都将生物对象的标准有效分辨率限制为 2nm（20Å）。虽然如此这大约比光学显微镜的分辨率要高 100 倍。此外，在最近这些年，随着一种叫做场发射枪的电子照射源的发展，电子显微镜的性能也得到了提高。这种非常亮且连续的电子源可充分提高能达到的分辨率。表 9-2 中列出了电子显微镜技术发展中的主要里程碑。

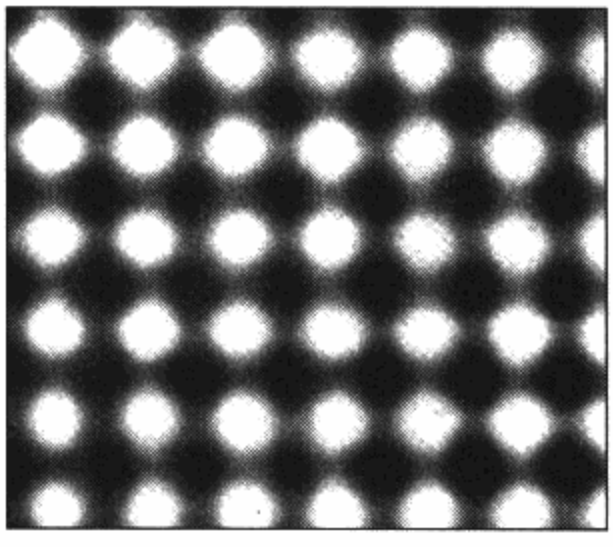


图 9-21　电子显微镜的分辨限度

这张关于金箔的透射电子显微照片中将晶体中原子纵列作为光斑展示出来。两列邻近的金原子之间的距离是 0.2nm（2Å）(经 Graham Hills 许可)

表 9-2　电子显微镜及其在细胞生物学中应用的发展过程中的主要事件

1897 年	Thomson 宣布带负电的颗粒的存在，稍候命名为电子
1924 年	de Broglie 提出运动的电子具有波动性
1926 年	Busch 证明了使用一圆柱形磁透镜聚焦电子束的可能性，奠定了电子光学的基础
1931 年	Ruska 和合作者制作了世界首台透射电子显微镜
1935 年	Knoll 论证了扫描电子显微镜的可行性；3 年后，Von Ardenne 建造了一台模型仪器
1939 年	Siemens 生产出第一台商业化的透射电子显微镜
1944 年	Williams 和 Wyckoff 引入金属投影技术
1945 年	Porter、Claude 和 Fullam 使用电子显微镜检测组织培养中的细胞，采用锇酸进行固定和染色
1948 年	Pease 和 Baker 用生物材料制备了超薄切片（0.1～0.2μm 厚）
1952 年	Palade、Porter 和 Sjöstrand 发展出固定和超薄切片方法，使得很多细胞内的结构首次得以观察到。这些技术首先被 Huxley 用来展示骨骼肌中蛋白质纤维的排列，证实了肌肉收缩时纤维滑动假说
1953 年	Porter 和 Blum 首次开发出普遍公认的超微切片机，将 Claude 和 Sjöstrand 引入的很多部件组合在一起
1956 年	Glauert 和合作者指出环氧树脂 Araldite 对于电子显微镜技术是一种很有效的包埋剂。5 年后，Luft 引进另一种环氧树脂 Epon
1957 年	Robertson 描绘了细胞膜的三层结构，这是第一次用电子显微镜观测到的

续表

1957 年	最初由 Steere 发展出的冷冻断裂技术，被 Moor 和 Mühlethaler 完善了。后来（1966 年），Branton 证明了冷冻断裂能够显示膜内部的结构
1959 年	Singer 使用偶联了铁蛋白的抗体用于电子显微镜下检测细胞内分子
1959 年	负染色技术由 Hall 在此 4 年前发明，Brenner 和 Horne 进一步发展了负染色技术，将这一通用的技术用于观察病毒、细菌和蛋白质纤维
1963 年	Sabatini、Bensch 和 Barrnett 引入戊二醛（通常接着用锇酸）作为电子显微镜中的固定剂
1965 年	剑桥仪器公司生产了第一台商业化的扫描电子显微镜
1968 年	de Rosier 和 Klug 给出了电子显微照片的三维结构重建技术
1975 年	Henderson 和 Unwin 从未染色样品的电子显微照片中，利用基于计算机的重构法，首次测定了一膜蛋白的结构
1979 年	Heuser、Reese 和合作者利用快速冷冻标本建立了高分辨率的深度蚀刻技术
1980 年	Dubochet 和合作者引入快速冷冻获得玻璃态冰薄膜用于高分辨率显微成像术
1997 年	Crowther、Fuller、Frank 和合作者利用单颗粒重建测定了病毒和核糖体的高分辨率（8～10Å）结构

透射电子显微镜（TEM）的整体设计和光学显微镜的相似，尽管它要大得多而且颠倒过来了（图 9-22）。照明源是能发出电子的一灯丝或阴极，置于一高约 2m 的圆柱的顶部。由于电子和空气分子碰撞后会被分散，所以空气必须从圆柱中抽出以形成真空。然后电子在附近阳极作用下从灯丝开始加速，通过一个小孔形成沿圆柱向下运动的电子束。沿圆柱每隔一段距离放置磁铁聚焦电子束，就像光学显微镜中用玻璃透镜聚焦光线一样。标本经过密封过渡仓被推入真空，进入电子束的轨道。和在光学显微成像中一样，标本通常要被染色，这种情况下使用电子稠密的材料染色，正如我们下一部分要了解的。一些电子通过标本时被染有电子稠密的材料散射；剩下的在摄影板材或磷光屏上被聚焦形成图像。因为散射的电子从电子束中丢失，图像中看上去暗的部分，是因为这部分电子通量减少了，显示的是标本的稠密区域。

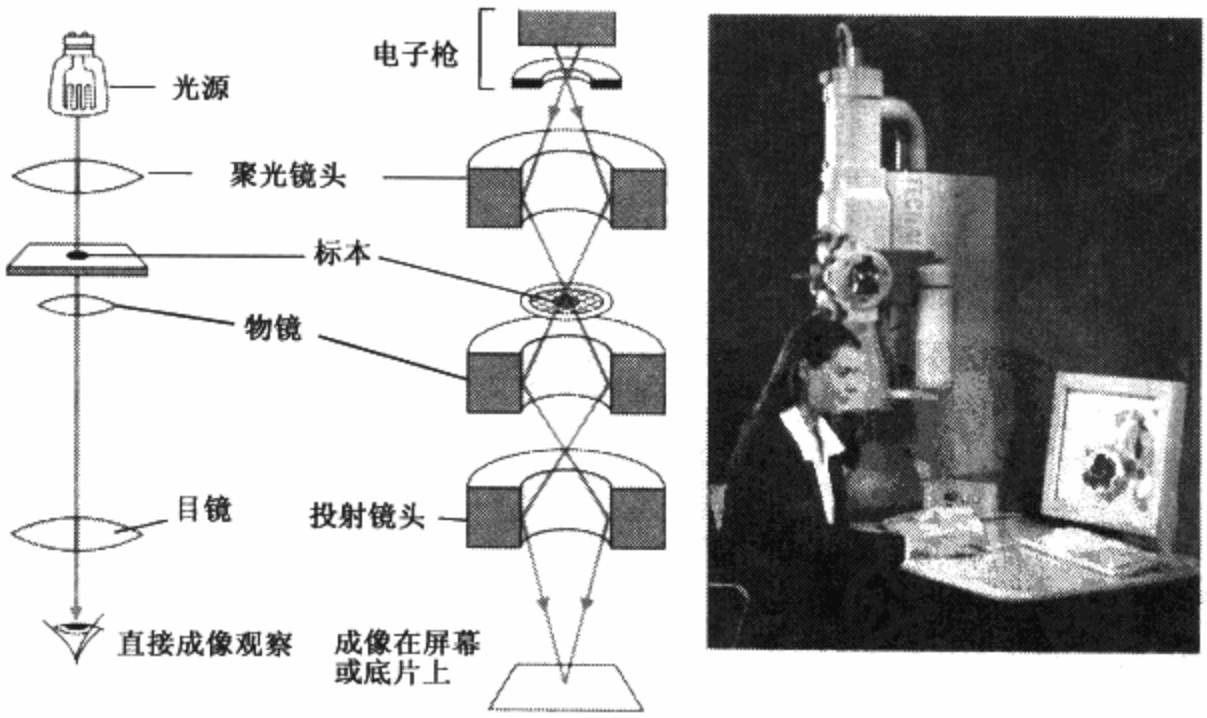


图 9-22　光学显微镜和透射电子显微镜的原理图

尽管光学显微镜中的透镜是由玻璃制成，电子显微镜中的透镜是磁性线圈。电子显微镜需要将标本放在真空中，但这些图强调了整体设计的相似性。插图显示的是使用中的透射电子显微镜

（照片经 FEI 有限公司许可）

用于电子显微镜的生物标本需要特别准备

在被用于生物材料的初期，电子显微镜展示了细胞中许多先前无法成像的结构。不过在产生这些发现之前，电子显微镜使用者不得不去发展出一些新的对组织包埋、切割和染色的过程。因为在电子显微镜中标本要暴露在非常高的真空下，因此不可能在活的、湿的状态下观察标本。组织通常被固定保存，最先用的是戊二醛，它使得蛋白质分子相互间共价交联，后来用四氧化锇（锇酸）固定，它结合到脂双层上，将脂双层像蛋白质一样稳定下来（图 9-23）。因为电子的穿透力非常有限，固定后的组织在观察前通常要被切割成非常薄的切片（50～100nm 厚，大概是一单细胞厚度的 1/200）。要做到这些，需要将样品脱水，然后用树脂单体渗透进去，树脂聚合后形成塑料固块；然后在专门的显微薄片切片机上用精细的玻璃或钻石刀切片。这些无水和其他挥发性溶剂的薄切片可以被放置在一个小圆金属网上用于显微镜观察（图 9-24）。

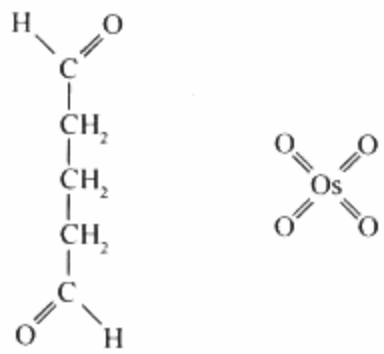


图 9-23　电子显微镜技术中常用的两种固定剂

戊二醛的两个反应性醛基能够和多种类型的分子交联，因而在分子之间形成共价键。四氧化锇能被许多有机物还原，并与之形成交联复合物。这对于固定细胞膜尤其有用，因为它会和许多脂肪酸中存在的C═C双键反应

准备用于观察的生物材料所需步骤从一开始就考验着使用电子显微镜的技术人员。我们如何才能确定那些经过固定、脱水、树脂包埋后的标本最终所显示的图像和那些活细胞中原本存在的精巧的含水的生物系统之间还有关系？对于这个问题现在最好的方法是快速冷冻。如果一个水相系统被足够快的冷冻到一足够低的温度，水和其中的其他成分没有时间重排结晶成冰。水被过度冷却转而形成一刚硬但非结晶的状态，就像一块玻璃，这称作玻璃态的冰。要得到这种状态，可以将标本猛地放在液氦冷却的光滑铜块上，或者将标本插入液态丙烷之类的制冷剂或将制冷剂喷在标本上，或者在高压下冷冻标本。

用特殊的冷冻标本台可以在电子显微镜下直接检测某些冷冻标本。其他情况下冷冻块可以被断裂暴露出内表面，或者将周围的冰升华暴露出外表面。不过我们通常希望观察薄切片，将切片染色在电子显微镜图像中产生适当反差。因此一个折中的方案是快速冷冻组织，然后用有机溶液替换掉玻璃态的水，最终将组织包埋到塑料树脂中，切片并染色。尽管技术上挺困难，但这个方法能将组织稳定保持在一非常接近原本的生活状态的环境中。

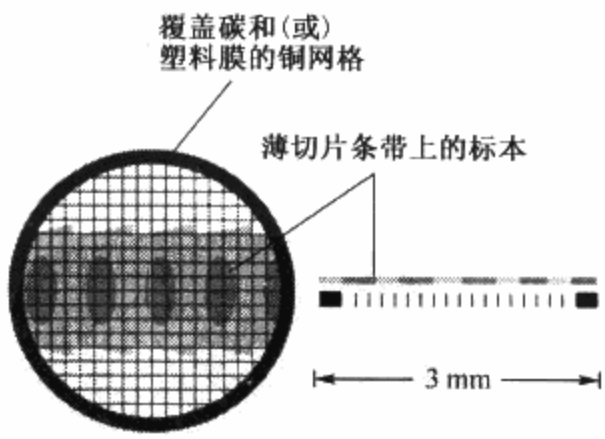


图 9-24　在 TEM 中支撑标本薄切片的铜网

电子显微镜中的反差依赖于标本中的原子序数：原子序数越高，电子被散射的越多，反差越强烈。生物组织由非常低的原子序数的原子（主要是碳、氧、氮、氢）组成。要使得它们能被看见，它们通常用铀或铅等重金属盐浸渍（在切片前或切片后）。细胞不同的组分根据在重金属盐中浸渍或“染色”程度的不同显示出不同程度的反差。例如，脂质在锇固定后往往被染黑，显露了细胞膜的位置（图 9-25）。

特定的大分子能用免疫金电子显微镜进行定位

我们已经了解抗体结合荧光显微镜技术如何用于特定的大分子的定位。**免疫金电子显微镜**方法作为一种类似方法，也能被用在电子显微镜中。通常的步骤是用一特定一抗孵育薄切片，然后

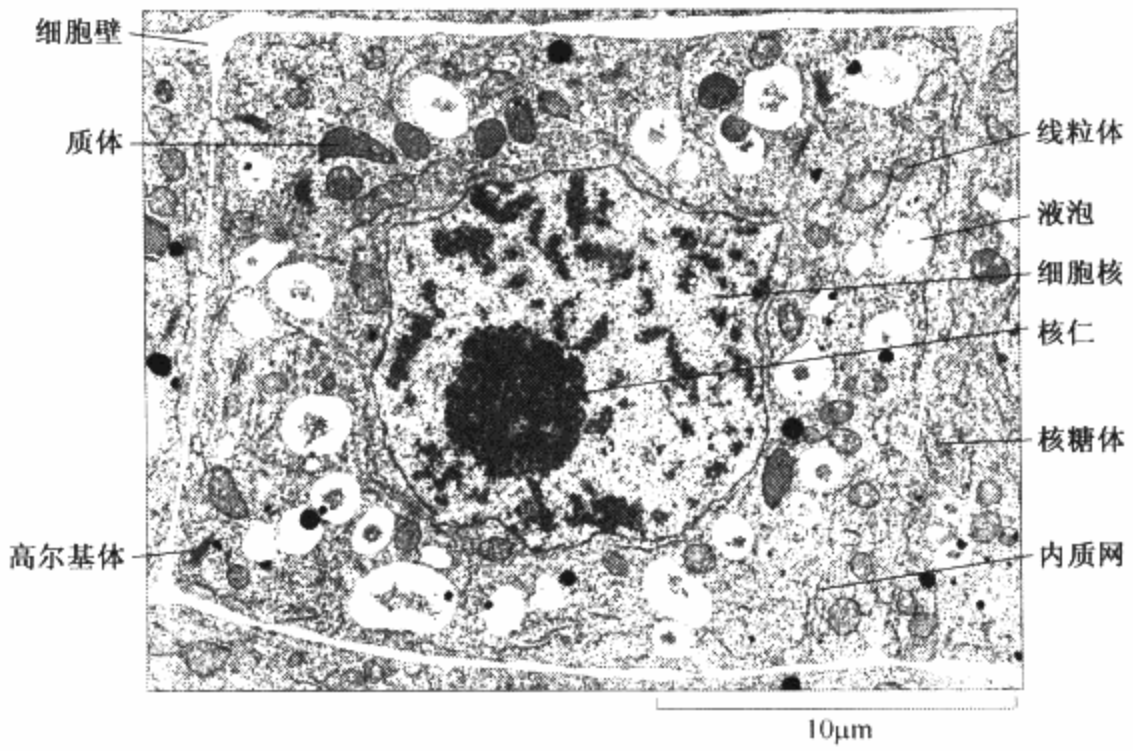


图 9-25　用锇和其他重金属离子染色的根尖细胞

在这张透射电子显微照片中，细胞壁、细胞核、液泡、线粒体、内质网、高尔基体和核糖体都清晰可见（经 Brain Gunning 许可）

和偶联了胶体金颗粒的二抗作用。金颗粒电子密度大，在电子显微镜中显示为一黑点（图 9-26）。

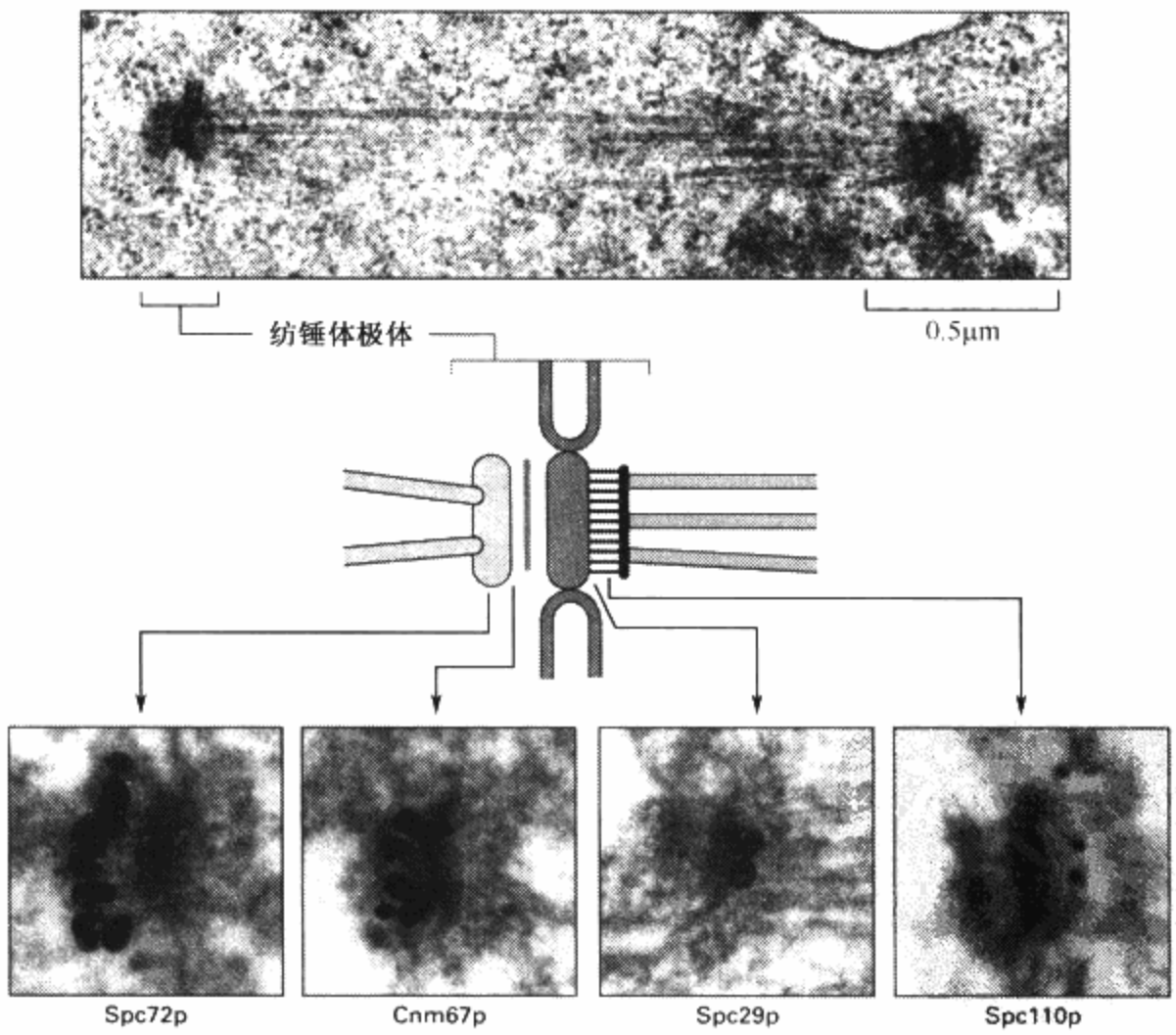


图 9-26　电子显微镜技术定位蛋白质

图中显示，运用免疫金电镜技术将酵母纺锤体极体中的 4 种成分定位于特殊部位。顶部照片所示是一个酵母有丝分裂纺锤体的超薄切片，显示纺锤体微管穿过细胞核，连接到包被在核膜中的纺锤体极体的末端。此照片下方显示的是单个纺锤体极体成分的简图。此技术中使用抗纺锤体 4 种不同蛋白质的抗体和胶体金（黑点），共同显示每种蛋白质在复合结构中的位置（经 John Kilmartin 许可）

使用透射电子显微镜通常得不到薄切片中细胞成分的三维排列，而且也很容易产生误解：例如，像微管之类的线型结构在切片中可能显示成点状物，而通过一个形状不规则固体的突出部分的切片，显示的可能是两个或更多个分散的物体。虽然能通过一系列的切片进行三维重构（图 9-27），但这仍然是冗长乏味的做法。

然而即使再薄的切片相对于电子显微镜的分辨率也要厚的多，因此即便是使用相应的方法也可能产生误解。电子显微镜使用非常小的光圈，这种光学设计能够得到大的景深，所以图像看上去相当于是不同深度结构的叠影。对于免疫金标记的更复杂的因素是抗体和胶体金颗粒渗入到包被树脂中；因此它们只能被用于检测切片表面的抗原。由于分布在切片内部的抗原分子不能被检测到，首先这意味着检测的灵敏度不高；其次，对于结构中是否包含抗原也可能得到错误的印象。这问题的一种解决方法是在塑胶包埋标本前，标记试剂还很容易靠近细胞和组织的时候，进行标记。直径大约 1nm 的非常小的金颗粒很适用于这种方法。像这么小的金颗粒在最终的切片中通常是不可见的，所以使用类似照相显影的化学方法，以 1nm 的金颗粒为核，用额外的银或金来包被。

用扫描电子显微镜获取表面图像

扫描电子显微镜（SEM）能够直接获取标本表面的三维结构图像。和透射电子显微镜相比，SEM 通常是更小、更简单也更便宜的设备。TEM 利用通过标本的电子来形成图像，而 SEM 利用标本表面反射或发出的电子成像。用于检测的标本需要固定、干燥以及在表面涂上一重金属薄层。此外，也能通过快速冷冻，然后将标本转到冷冻标本镜台上用显微镜直接观察。通常完整的植物或小动物几乎不需要什么准备过程就可以直接放到显微镜中（图 9-28）。然后用一束非常窄的电子束扫描上述方法准备的标本。测量第一束电子轰击金属表面时散射

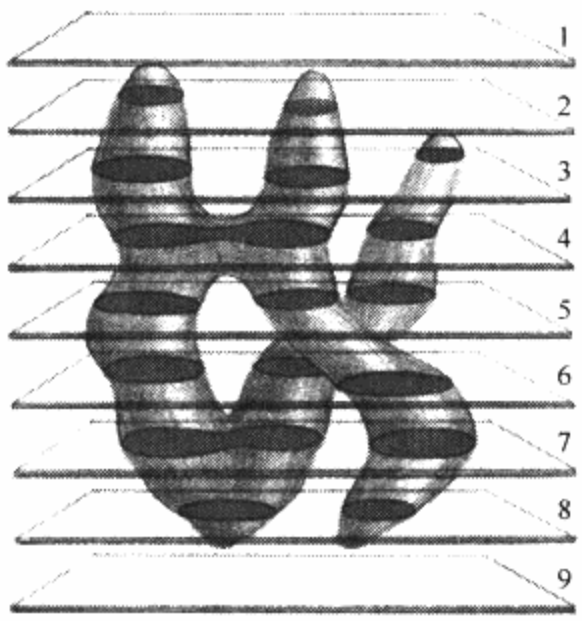


图 9-27　一系列切片的三维重构

单个薄切片有时会造成令人误解的印象。在这个例子中，包含有一个分岔的线粒体的细胞的切片大部分看上去含有二三个分离的线粒体。而且切片 4 和 7 可能会被认为显示了不同处理下的线粒体。然而，从这一系列切片中可以重构出真实的三维形状

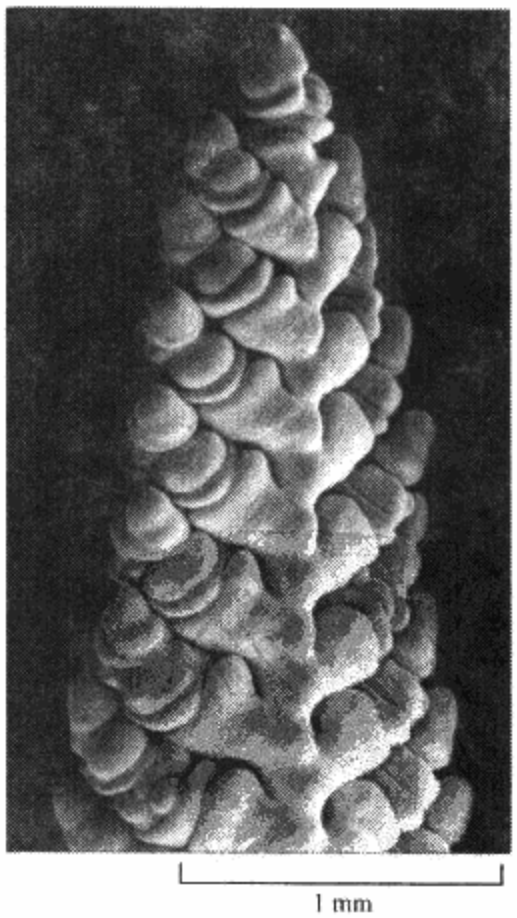


图 9-28　一个发育中的小麦花或穗

这个精致的花穗被快速冷冻，用一薄金属层包被，在冷冻状态下用 SEM 检测。这一低倍的显微照片，显示了 SEM 的大景深（经 Kim Findlay 许可）

或发出的电子的量，并和另一束完全同步的电子束进行对比，然后成像在电视荧光屏上。使用这种方法，可以得到高度放大的表面图像（图 9-29）。

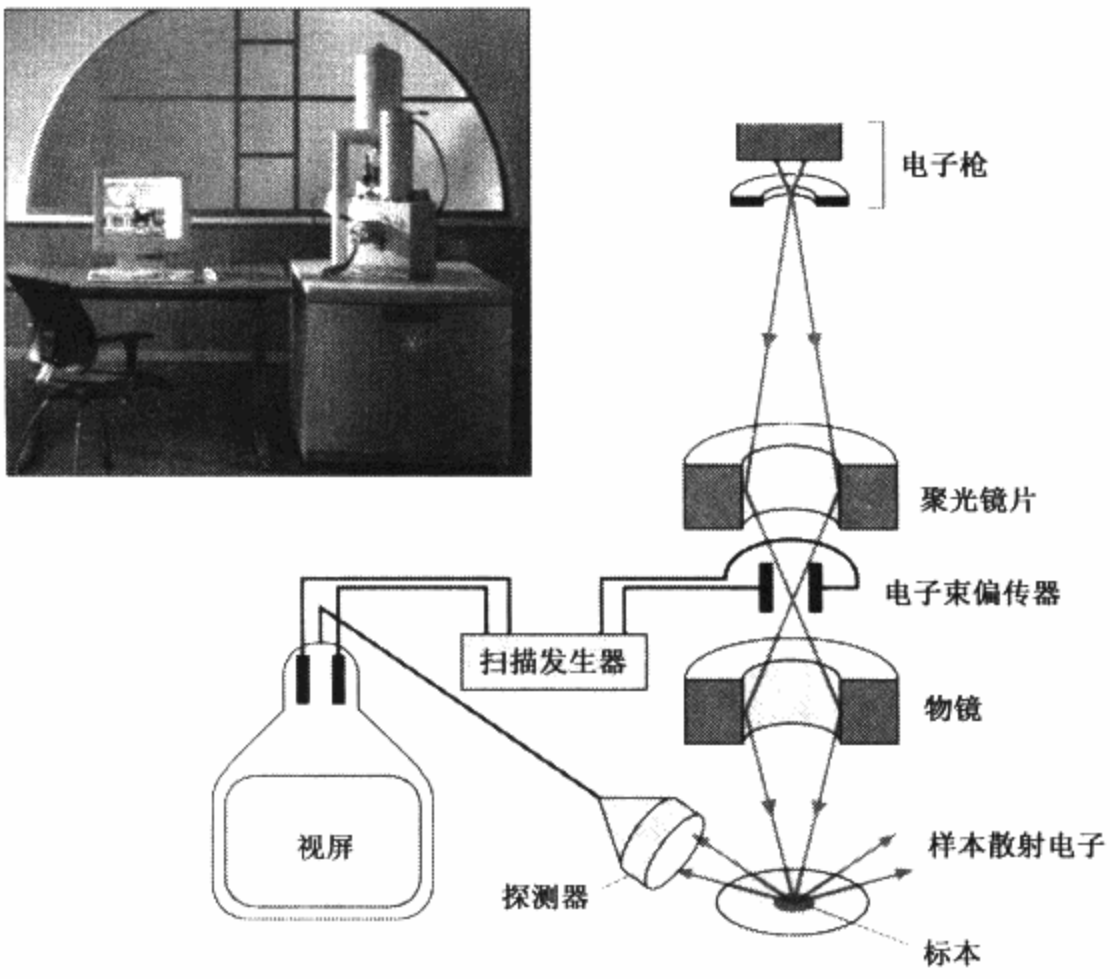


图 9-29　扫描电子显微镜

在一个 SEM 中，样品被一束聚焦在样品表面的电子束扫描，其中一个电磁卷板起到透镜作用电子束轰击表面连续的点时散射和发出的电子数量通过探测器测量，这些电子的数量被用来控制显示屏所放大的图像中连续点的亮度。SEM 可以获得三维物体大景深的突出的图像，图像的分辨率根据所用设备在 3nm 到 20nm 之间（照片经 FEI Company Ltd 许可）

SEM 技术提供了大的景深；此外，由于电子散射的数量依赖于电子束和表面之间的角度，图像通过高亮和阴影给出了三维外观（图 9-28 和图 9-30）。但是只有表面特征才能被检测出来，而且多数 SEM 能达到的分辨率不是太高（大约 10nm，有效放大倍率为 20 000 倍）。因此，这种技术通常被用于研究全细胞和组织，而不是亚细胞器。不过最近一种使用亮连续场发射枪作为电子源的具有非常高分辨率的 SEM 已经发展起来了。这类 SEM 产生的图像在分辨率上可以和 TEM 产生的图像媲美（图 9-31）。

在透射电子显微镜的高分辨率下通过金属阴影来检测表面特征

TEM 也能被用来研究标本的表面，而且一般来说比使用 SEM 的分辨率要高，使用 TEM 能够观测到独立的大分子。就像在扫描电子显微技术中一样，干燥的标本表面也被喷上铂之类的重金属薄层。金属呈斜角喷溅上去，以至于最后沉淀下来后某些地方的金属层会比其他地方的要厚些，这个过程被称为投影法，因为这种投影效果产生的阴影能够给图像以三维外观。

某些用这种方法包被的标本足够的薄且小，电子束能够直接穿过它们。对于单个分子、病毒或者细胞壁，在投影法之前，先要被干燥到一张平的支撑膜上，这种膜对电子有相应的通透性，如用碳或塑料。对于较厚的标本，在投影法之后，必须要将细胞中的有机物都溶解掉，只留下标本表面的薄金属复制层。在用碳膜来加固这复制层，然后放在架子上，按照常

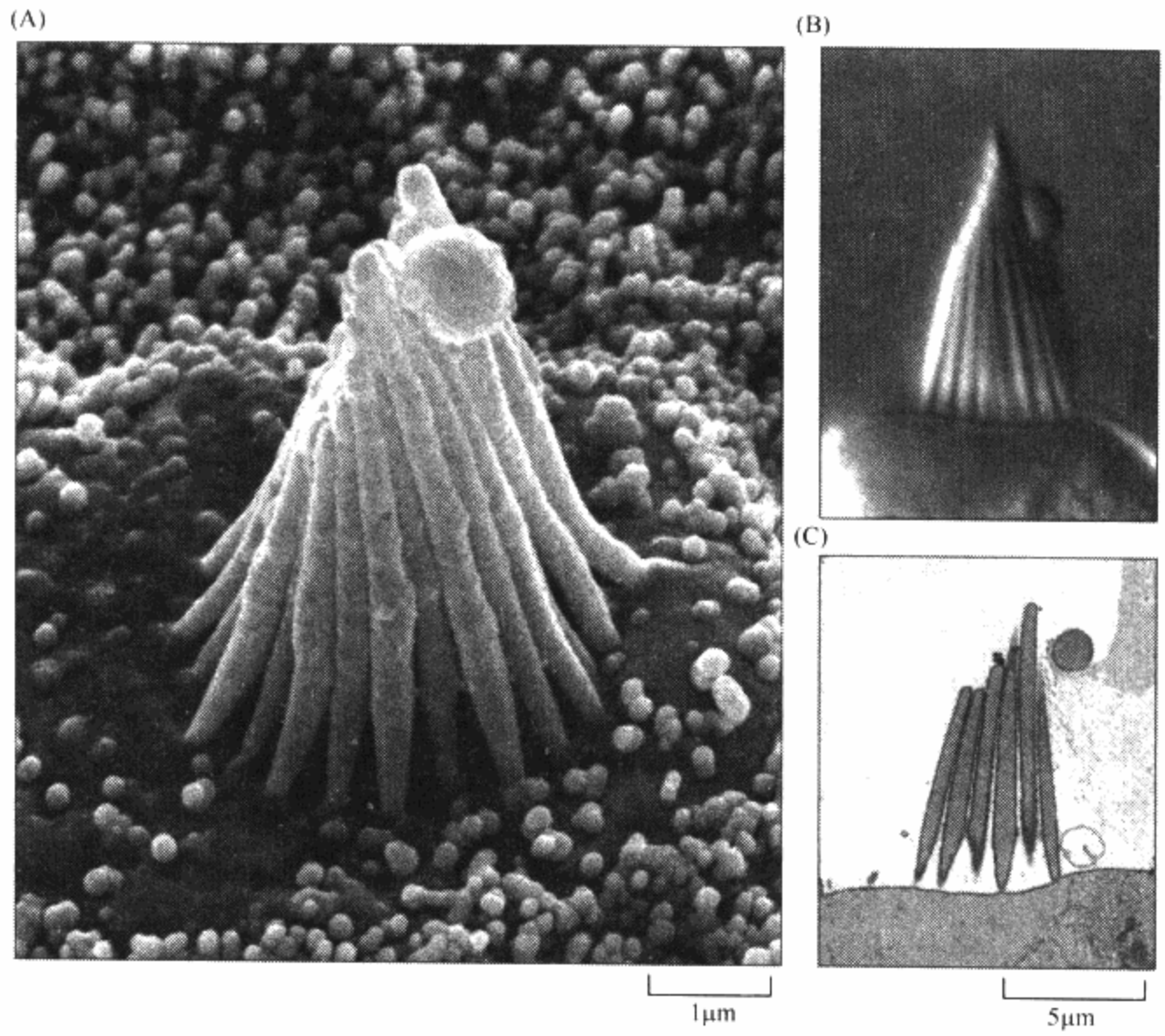


图 9-30　扫描电子显微成像术

（A）牛蛙内耳中的绒毛细胞突出的纤毛的扫描电子显微照片。为了对比，相同的结构也通过（B）微分干涉相差光学显微成像和（C）薄层透射电子显微成像来显示（经 Richard Jacobs 和 James Hudspeth 许可）

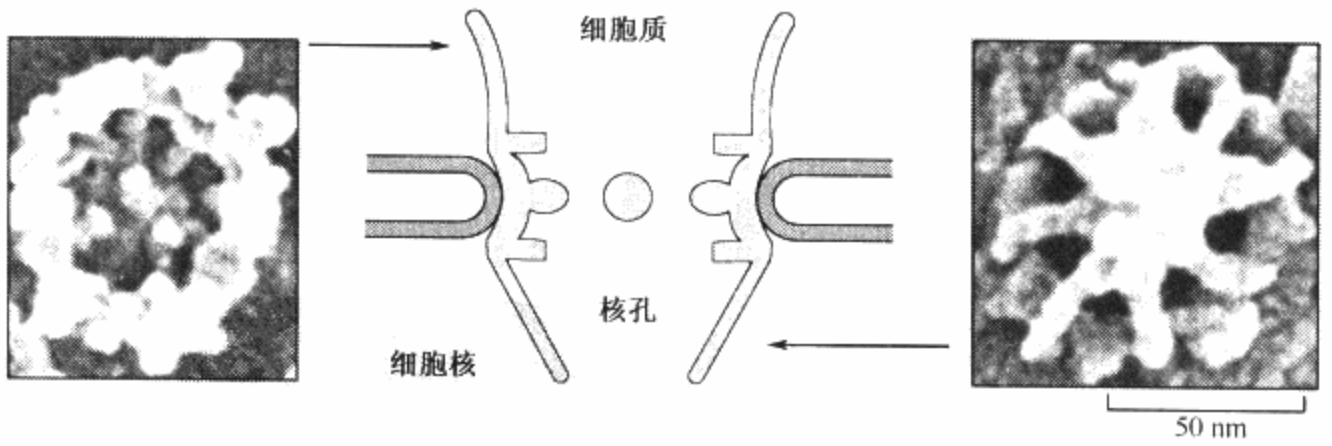


图 9-31　核孔

快速冷冻的核膜在高分辨率的 SEM 下成像，该 SEM 配有一个场发射枪作为电子源。核孔每侧的这些图像体现了 SEM 分辨率的极限，这可以和图 12-10 比较（经 Martin Goldberg 和 Terry Allen 许可）

规方法用透射电子显微镜进行检测（图 9-32）。

冰冻断裂和冰冻蚀刻电子显微镜方法获取细胞内部的面层图像

冰冻断裂电子显微镜方法提供了一种观测细胞膜内部的方法。细胞被冷冻后用一刀片破开。断裂面通常经过脂双层的疏水中心，从而露出细胞膜的内部。获得的断裂面用铂投影，

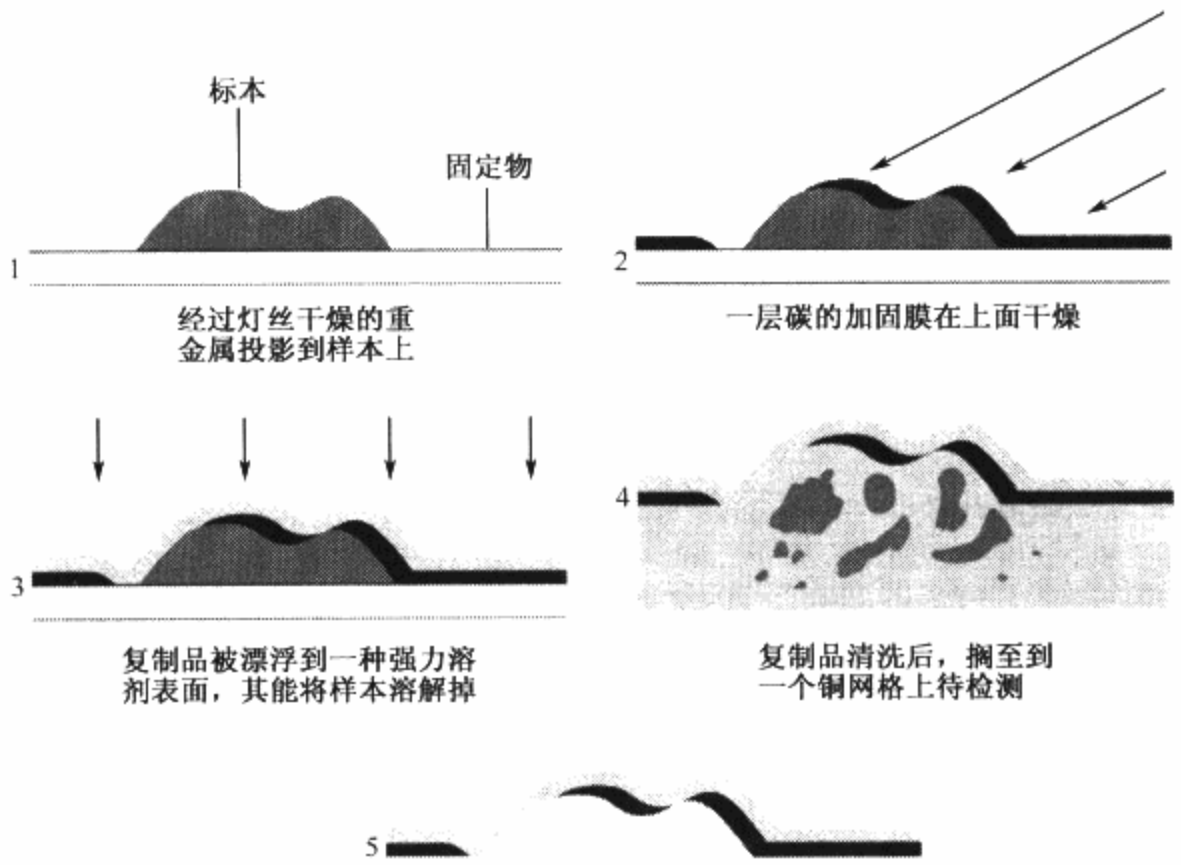


图 9-32　标本表面金属投影复制层的准备

注意金属的厚度反映了原始标本的表面轮廓

有机物被溶解掉，复制层浮起后用电子显微镜观测（图 9-32）。膜内颗粒使得这种复制层上呈现很多小凸起，象征着大的跨膜蛋白。这种技术提供了一方便生动的显现膜平面上蛋白质分布的方法（图 9-33）。

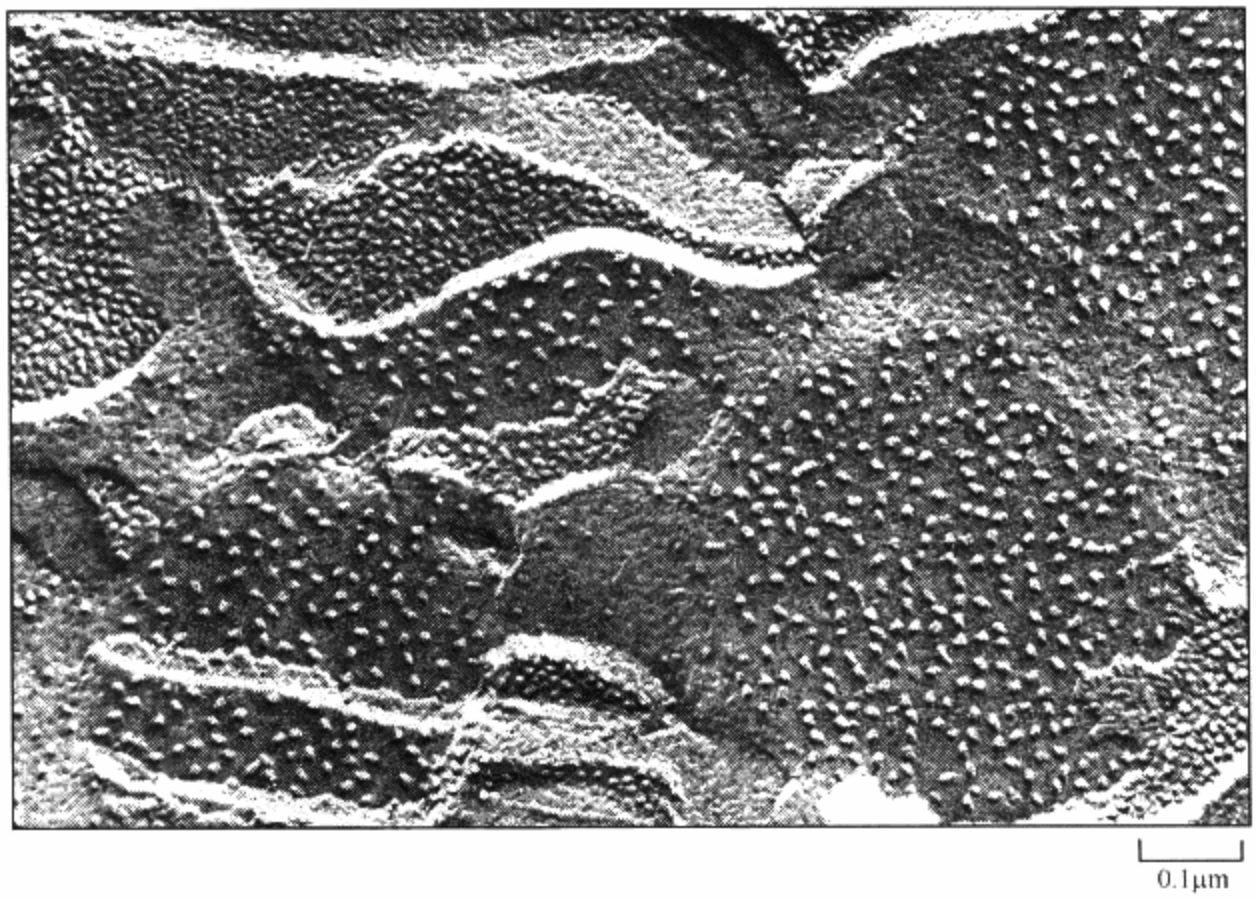


图 9-33　植物叶绿体中的类囊体膜

此图为冰冻断裂电子显微镜照片，图中显示，执行光合作用的类囊体的膜是堆积成多层（图 14-34）。断裂面从一层跨至另一层，途经每个脂双层的中部，并暴露出双分子层中的跨膜蛋白，这些蛋白质的体积很大，因而能在铂复制层（platinum replica）中以膜内颗粒的形式投影并显示出来。膜中所见的最大颗粒是完整的光合系统Ⅱ——一个由多种蛋白质形成的复合体（经 L. A. Staehelin 许可）

另外一种方法是冰冻蚀刻显微镜方法，这种方法既可用于检测细胞外表也能检测细胞内部。这种技术中，冰冻样品像上述过程一样用刀片骤然撕开。然后稍微升温，在真空中冰升华（这过程也称作冻干），冰层低于细胞。通过蚀刻过程细胞暴露出来的部分然后再用前面提到的投影法制作一个铂的复制层。这种技术能够暴露出细胞的内部结构，并且能非常清晰地显示出这些结构的三维构造（图 9-34）。

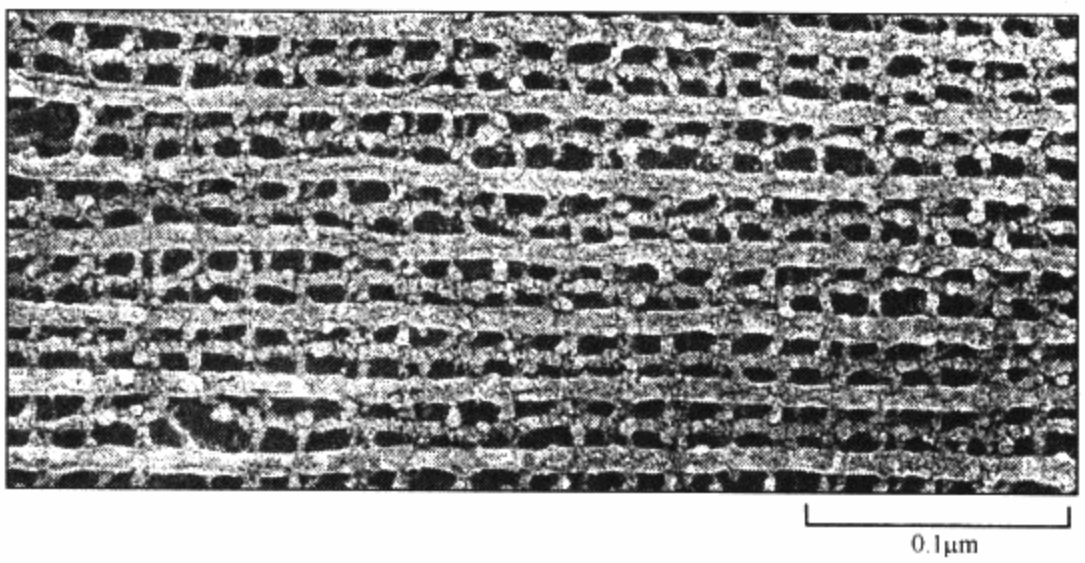


图 9-34　在昆虫肌肉中蛋白质纤维的规则排列

要获得这一图像，肌肉细胞要被快速冷冻到液氮的温度，从细胞质断裂，经历深度蚀刻。然后准备一个金属投影的复制层，在高分辨率下检测（经 Roger Cooke 和 John Heuser 许可）

负染色和低温电子显微成像术能在高分辨率下观察到大分子

尽管使用重金属投影法提供对比后，能够容易地用电子显微镜观察 DNA 或大蛋白质等独立的大分子，但是使用负染色能够更好地观察到细微结构。负染色技术中，使用乙酸双氧铀之类的重金属盐溶液浸泡由碳薄膜支持的分子。样品干燥后，除了大分子所在的地方，其他地方都覆盖了一薄层重金属盐。因为和周围的重金属染料相比，大分子更容易让电子通过，因此可以得到反转的或负的图像。负染色对于观察大分子的聚合体尤其有用，如可用于观察病毒、核糖体和蛋白质丝的亚组分结构（图 9-35）。

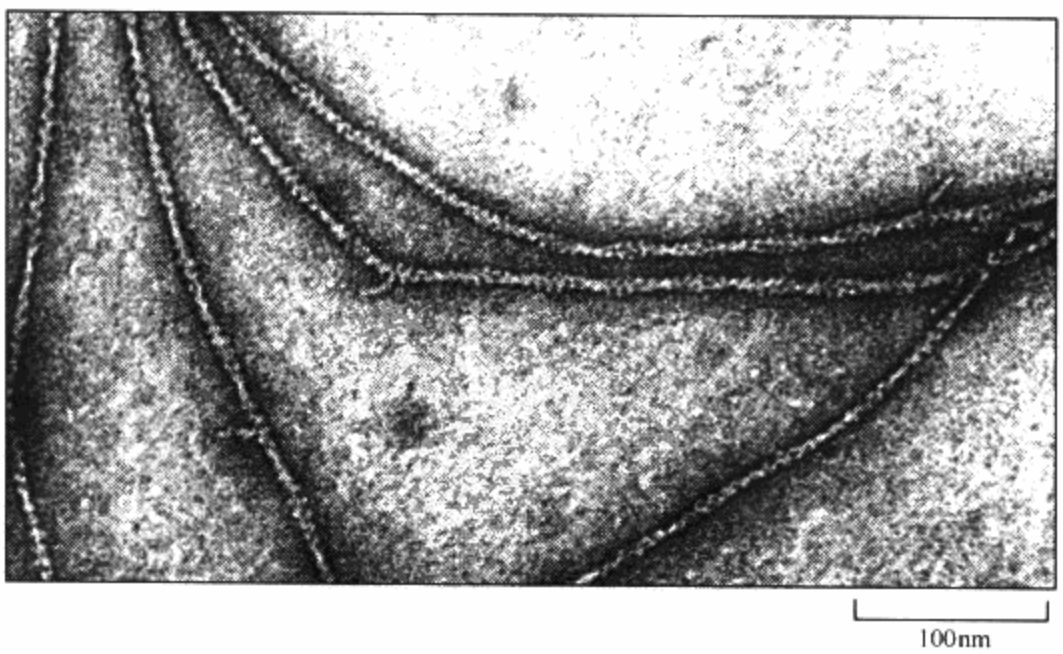


图 9-35　肌动蛋白丝负染色

在这张透射电子显微镜照片中，每根肌丝直径大约为 8nm，进一步观察发现肌丝看上去是由球形肌动蛋白分子形成的螺旋链所组成的（经 Roger Craig 许可）

投影法和负染色都能提供大分子聚合体高对比度的表面图像，不过这两种技术的分辨率都受限于投影或染色中所使用的最小金属颗粒的尺寸。新近的方法能够直接在高分辨率下观察病毒之类的内部形态的三维结构。在这种被称作**低温电子显微成像术**的技术中，关键是要快速冷冻形成玻璃态冰晶。病毒或大分子复合物的水悬液在电镜栅上形成一层非常薄的膜（大约 100nm）。然后将标本放进冷冻液中快速冷冻。在显微镜的真空器中用一特殊的样品台使含水标本保持在－160℃，不需要固定、染色和干燥就可以直接观察。和负染色方法不同的是，水合低温电子显微镜成像术能直接获取大分子本身的图像。不过，正如我们接下来要讨论的，要想尽可能的获得结构信息，需要特定的图像处理技术。

多重图像能组合到一起提高分辨率

无论是从电子显微镜还是从光学显微镜获得的图像，都是由粒子（电子或光子）撞击某类检测器得到的。不过这些粒子受量子力学的约束，因此预计到达检测器的粒子数量只有统计学意义。在粒子数量非常多的情况下，反映标本图像的粒子在检测器的分布是可以精确测定的。但是在粒子数量较少的情况下，在每个区域被检测到的粒子数量呈统计上的波动，导致底图的模糊。这种扰乱标本底图的随机波动归于噪声。噪声对于未染色大分子的电子显微成像方法来说是特别严重的问题，而且这对于低光照水平下的光学显微成像也很重要。一个蛋白质分子要不被破坏所能承受的电子照射量每次每平方纳米不超过 10 个电子，这种照射量低于原子分辨清晰成像所需。

上述问题的解决方案是获取许多同样分子的图像，很可能需要上万张单独的图像，然后将这些图像合并起来得到一平均图像，来揭示原来图像中被噪声掩盖的结构细节。不过在这些单独的图像合并之前，它们必须彼此对齐。有时可能需要促使蛋白质或复合物结晶，每个分子能按规则点阵同向排列。这种情况下，图像对齐的问题很容易解决，用这种电子晶体学方法已经测定了原子分辨率下的几种蛋白质结构。不过原则上结晶阵列不是绝对必需的。借助计算机，任意分布的分子的图像可以被加工并合并进行高分辨率重构。我们接下来要说明这一点。

来源于不同方向的图像能被组合到一起进行三维重构

记录来自电子显微镜的图像的探测器展现了二维图像。因为显微镜的大景深，三维标本的所有部分都是清晰的，因此得到的图像是沿观察方向的结构的投影。如果我们从许多不同的方向观察同一标本，第三维上丢失的信息可以被恢复。用于这项技术的计算方法在 20 世纪 60 年代被设计出来，现在已经被广泛用于医学计算机体层摄影（CT）扫描。在 CT 扫描中，成像设备在患者周围移动获取不同的图像。在**电子显微镜（EM）断层扫描成像术**中，显微镜中的标本台被倾斜，这也能达到同样的效果。使用这种方法，将显微镜视野中许多同样分子的一系列图像合并，可以得到某一选定方向上的三维重构图像。每一图像个别来看时噪声非常大，但合并三维的图像并取一平均值，噪声能被大量消除，得到分子结构的清晰图像。

电子显微镜断层扫描成像术现在被广泛用于测定结晶或非结晶的标本中的分子结构以及细胞或细胞器的薄切片等大的目标。对于那些具有内在对称的结构，像螺旋或二十面体病毒，这是一项特别成功的技术，因为它使得图像调整的任务更简单也更精确。图 9-36 显示了高分辨率下利用图像组合技术测定的一个二十面体病毒的结构，图 9-37 显示了同样方法测定的核糖体的结构。

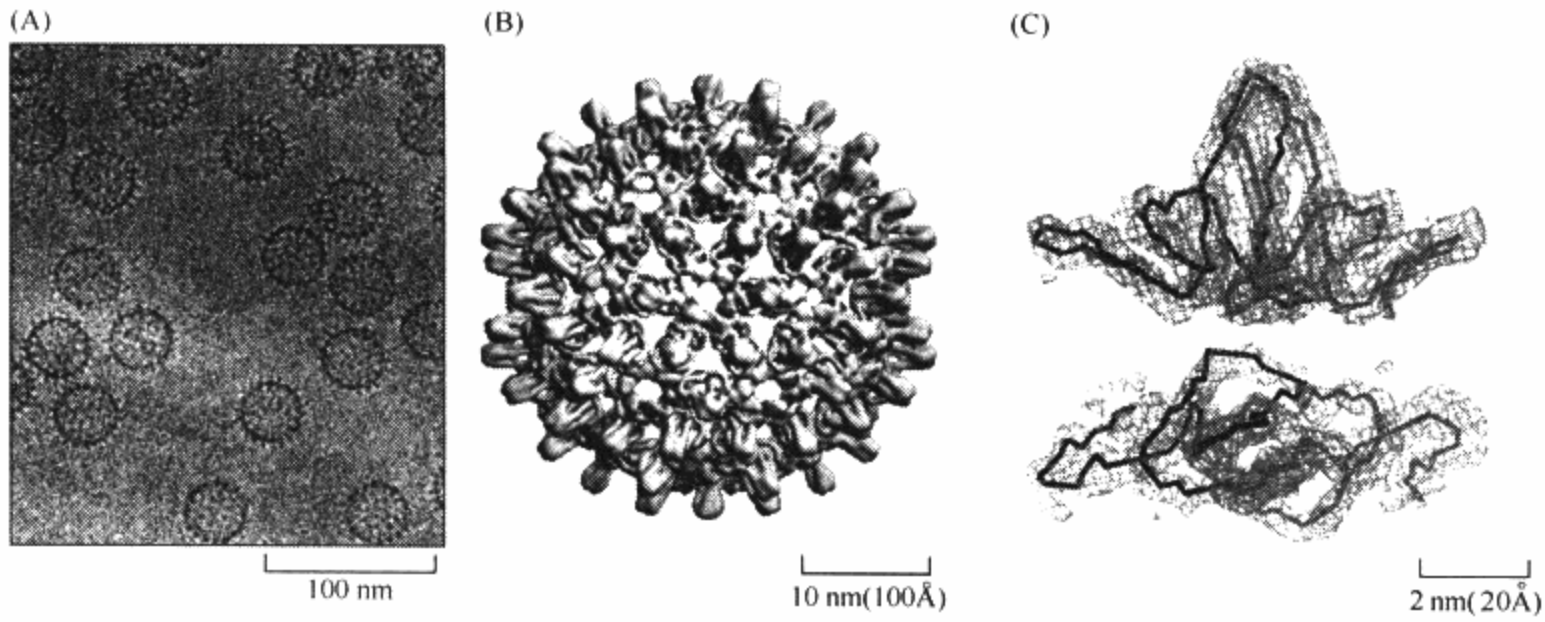


图 9-36　电子显微镜断层扫描成像术

乙肝病毒球形蛋白质外壳被保存在一薄层冰中（A）并在 TEM 下成像。通过电子显微镜断层扫描成像方法上千个这种颗粒被组合到一起得到这个二十面体颗粒的三维图像，显示在（B）中。（C）单个蛋白二聚体的两幅图像，这些二聚体形成了外壳表面的那些穗，这两幅图像显示重构的分辨率（7.4 Å）足够分辨多肽链全部的折叠［（A）经 B. Böttcher，S. A. Wynne 和 R. A. Crowther 许可；（B）和（C）摘自 B. Böttcher，S. A. Wynne and R. A. Crowther，*Nature* 386：88～91，1997. © Macmillan Magazines Ltd］

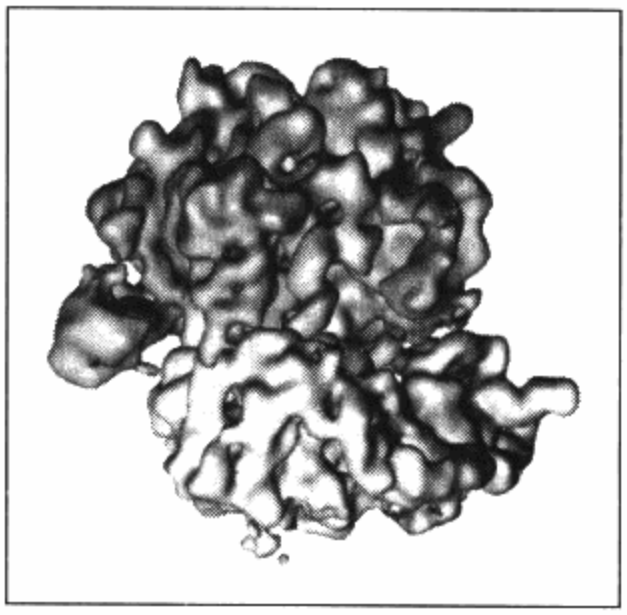


图 9-37　用电子显微镜断层成像术测得 *E. coli* 70S 核糖体的三维结构

小亚单位被染成黄色，大亚单位被染成蓝色。分辨率为 11.2Å（摘自 I. S. Gabashvili et al.，*Cell* 100：537～549，2000. @Elsevier）

用晶体的阵列，电子显微成像术达到了 0.3nm 的分辨率，这足以观察到蛋白质的内在原子排列，也足以和 X 射线晶体学的分辨率相匹敌。对于单微粒的重构，分辨限度大约是 0.8nm，这足以识别蛋白质亚组分和结构域以及有些蛋白质的二级结构。尽管作为大分子结构测定方法，电子显微成像术不太可能替代 X 射线晶体学（讨论见第 8 章），但它有一些非常明显的优势。首先，它并不完全需要结晶标本。其次，对于由于太大或不稳定以至于得不到好的结晶的复杂结构，它也能处理。电子显微成像术提供了单分子尺度和全细胞尺度之间的桥梁。

小结

许多光学显微技术可用于观测细胞。固定和染色后的细胞能用传统的光学显微镜来研究，偶联了荧光染料的抗体可用在荧光显微镜下定位细胞中特定的分子。活细胞可用相差显微镜、微分干涉相差显微镜、暗视场显微镜或明视场显微镜来观察。电子成像处理技术推动了光学显微镜的各种形式，这能增强图像的灵敏度、改善图像质量。共聚焦显微镜和图像去卷积都提供了更薄的光学截面并可用于三维图像的重构。

要检测细胞中膜和细胞器的细微结构需要更高的分辨率，透射电子显微镜能达到这一点。特定的大分子可以通过偶联了胶体金的抗体来进行定位。细胞和组织表面的三维图像可以通过扫描电子显微镜获取。通过重金属投影或负染色勾画获得的独立的大分子的形状，也能容易地用电子显微镜检测出来。使用计算的方法，通过一种称为电子显微镜断层扫描的技术，

能够将不同方向获得的多重图片进行大分子和分子复合物的细节重构。

观察活细胞中的分子

在细胞生活周期，即使最稳定的细胞结构也要经过装配、分解和重组。当细胞引导自身行为以及对环境应答时，其他一些在分子尺度上很大的结构也会快速变化、移动和重组。复杂的、高度组织化的分子机器在细胞中移动组分，控制组分进出细胞核、从一个细胞器到另一个细胞器以及进出细胞。

这部分我们将描述一些用于活细胞中研究动态过程的方法。所有的成像都需要使用某种形式的放射线，光线是对生命系统破坏最少的射线之一。已经发展出多种技术在显微镜中观察活细胞的特定分子。这其中大多数方法基于荧光标记或指示剂的使用。小到 Ca^{2+} 或 H^+ 之类的无机离子，大到特定蛋白质、RNA 或 DNA 等大分子都可以用这种方法清楚地成像。不过光学显微成像术并不是解决这一问题的唯一方法，显微镜也不是唯一必需的设备。

使用发光指示剂度量细胞内离子浓度的快速变化

研究单个活细胞化学的一种方法是将一纤细的、离子敏感的玻璃**微电极**直接穿过细胞质膜插到细胞内部。这项技术通常用于测量 H^+、Na^+、K^+、Cl^- 和 Ca^{2+} 等常见无机离子在细胞内的浓度。但是离子敏感微电极只能显示细胞中某一点的离子浓度，对于 Ca^{2+} 之类的非常低浓度的离子响应慢且有些不稳定。因此，这些微电极并不能完美地记录下细胞质中钙离子浓度快速和瞬时的变化，而钙离子在调控细胞对胞外信号的应答起重要作用。这种变化可以用离子灵敏指示剂来分析，**离子灵敏指示剂**发出的光反映了离子浓度。这些指示剂其中一些是发光的（自然发光），另一些是荧光的（感光后发光）。

Aequorin（发光蛋白质）是一种从水母中分离得到的发光蛋白；它在钙离子存在时发出光，并且当钙离子浓度在 0.5～10μm 范围内变化时做出反应。例如，将发光蛋白质显微注射到一个卵中，当卵受精时，在局部会有自由钙离子突然释放到细胞质中，此时发光蛋白质做出应答瞬间发光（图 9-38）。发光蛋白质也在植物和其他生物中转基因表达，这也提供了一种无需显微注射就可监测这些生物细胞中钙离子的方法，而显微注射可不算一种容易的操作。

发光蛋白质之类的生物发光分子发出微量的光，每个指示剂分子最多只能发出几个光子，这很难测量。荧光指示剂每个分子能产生更多的光子，因此它们更容易被测量并且能给出更好的空间分辨率。牢固结合了钙离子的荧光钙离子指示剂和没有结合钙离子的指示剂在激发光或发射光的波长上有轻微不同。通过测量在两种激发光或发射光波长下荧光强度的比率，钙离子结合的指示剂和未结合钙离子的指示剂的浓度比率能够被测定，这样就提供了一种测定自由钙离子的浓度的灵敏方法。这种类型的指示剂被广泛用于荧光显微镜下实时监测细胞内不同部分钙离子浓度的变化（图 9-39）。

相似的荧光指示剂可用于测量其他离子。例如，某些被用于测量氢离子和细胞内的 pH。这类荧光指示剂中有些能够通过扩散进入细胞，因此不需要显微注射；这也使得在荧光显微镜下同时检测大量的单个细胞成为可能。新型的指示剂和新的图像处理方法联合起来，能得到许多类似的快速和准确的分析细胞内多种小分子浓度变化的方法。

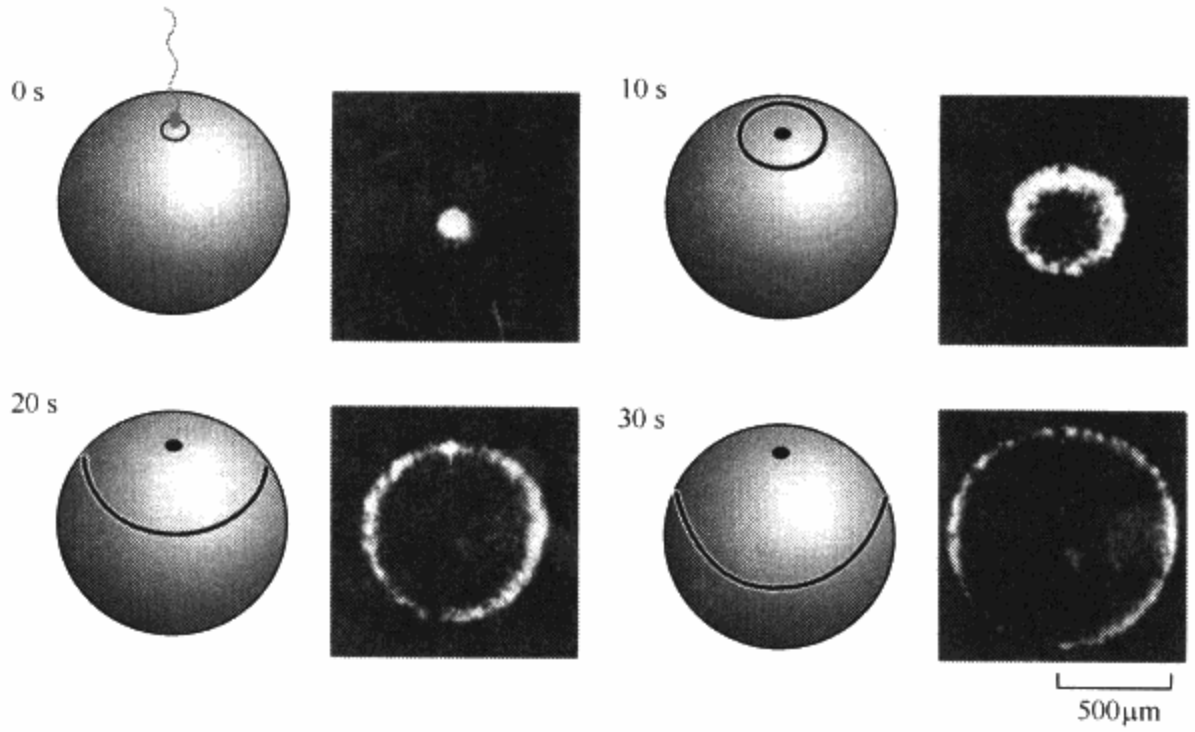


图 9-38　一种发光蛋白 Aequorin

Aequorin 蛋白在有游离 Ca^{2+} 存在时发光。图中所示为一个被注射了 Aequorin 蛋白的锵鱼卵，这种蛋白质在全部细胞质中扩散，然后这个卵和精子受精，在图像增强剂的辅助下成像。这 4 张俯视照片是在精子进入位点的上方每隔 10s 拍摄的，显示了从质膜下的内部储备中释放游离 Ca^{2+} 进入细胞质的波。穿过这个卵的波起始于精子进入的位点，如左边的简图所显示（摘自 J. C. Gilkey, L. F. Jaffe, E. B. Ridgway and G. T. Reynolds, *J. Cell Biol.* 76: 448～466, 1978. © The Rockefeller University Press）

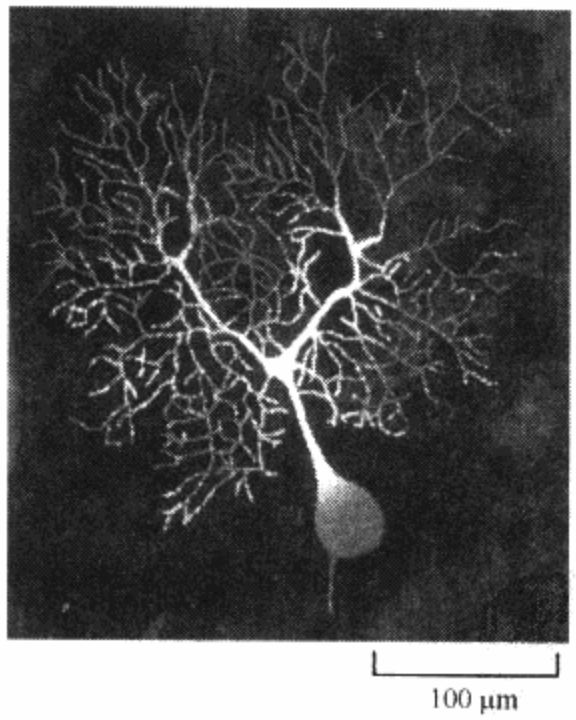


图 9-39　使用荧光指示剂观察细胞内的 Ca^{2+} 浓度

小脑中的一个浦肯野细胞树突分支从其他神经元接收到 100 000 个以上的突触。照片底部显示细胞的输出信号离开细胞体沿着轴突传递。单个浦肯野细胞（来自豚鼠）的胞内 Ca^{2+} 浓度的图像通过一个低照度相机拍摄，其中采用了对 Ca^{2+} 敏感的荧光指示剂 fura-2。通过不同的颜色表现出自由 Ca^{2+} 的浓度，红色代表最高浓度，蓝色代表最低浓度。最高的 Ca^{2+} 水平呈现在成千个的树突中（经 D. W. Tank, J. A. Connor, M. Sugimori 和 R. R. Llinas 许可）

几种将不透膜的分子导入细胞的方法

无论是识别细胞内蛋白质的抗体，还是荧光探针标记的普通细胞蛋白质，或者是影响细胞行为的分子，能将这些不透膜的分子导入活细胞中通常很有用。其中一种方法是使用微量吸管将分子显微注射入细胞。有一种特别有用的技术被称作**荧光模拟细胞化学**，在此方法中纯化好的蛋白质被偶联上荧光染料并被显微注射进入细胞。当细胞生长和分裂的过程中，使用荧光显微镜可以追踪到该注射的蛋白质。如若将微管蛋白（微管

的亚组分）用红色荧光染料标记，将能在活细胞中实时追踪微观动态（图 18-21 和图 16-12）。

抗体可以被显微注射入细胞，去阻碍那些抗体所识别的分子的功能。例如，将抗肌浆球蛋白Ⅱ型的抗体注射入海胆卵中，将阻止卵细胞一分为二，即使细胞核分裂已经照常发生。这一观察证明了细胞分裂中肌浆球蛋白在分裂细胞质的收缩过程中担任的必需角色，但对于细胞核的分裂它并不是必需的。

虽然已经广泛使用，显微注射要求每个细胞单独被注射；因此一次至多只能研究几百个细胞。其他有些方法可允许大群细胞同时被透过。例如，其中有一种能部分地破坏细胞质膜的结构，提高膜通透性；这通常可以通过使用一强力电击或低浓度的去垢剂等化学物质来实现。这种电学技术可在质膜中造成许多大孔而不伤害细胞内的膜。这些孔可以保持开放状态几分钟到几小时，这依赖于细胞类型和电击大小，这些孔允许大分子快速进出细胞溶质。这种电穿孔的方法作为将 DNA 分子导入细胞的方法对于分子遗传学也有价值。用一有限的处理，这些细胞中的大部分修复了它们的细胞质膜存活下来。

第三种将大分子导入细胞的方法是使用包含有大分子的微囊和细胞质膜融合。要将新基因导入细胞核，可以将包被了 DNA 的金颗粒高速射入细胞。这些方法都已经在细胞生物学中广泛使用，如图 9-40 所示。

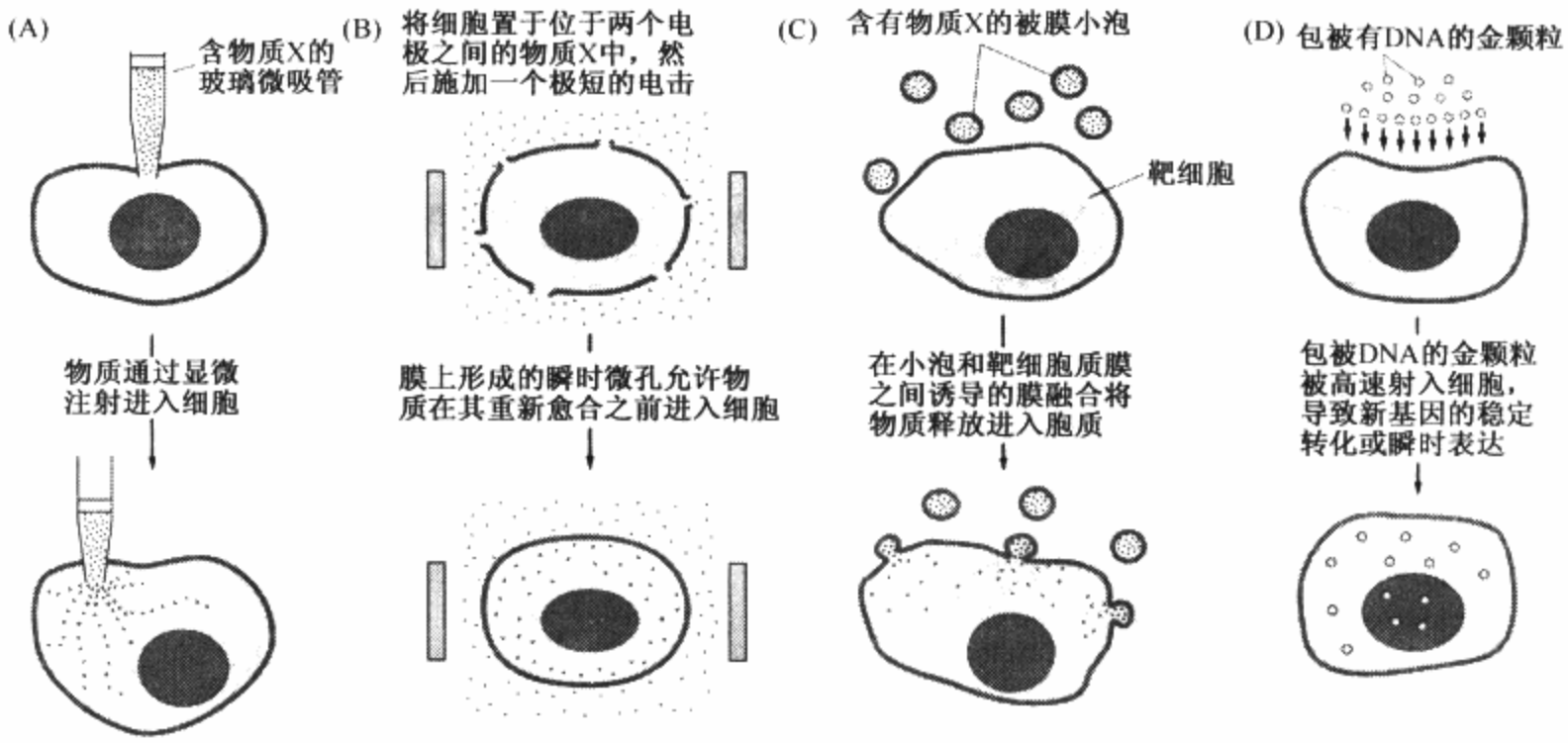


图 9-40　将不透膜的物质导入细胞的方法

（A）依靠施加压力通过一个微吸管将物质注射到细胞中，如果物质带电，还可通过施加电压驱使物质形成离子流进入细胞［一种被称作离子透入法（iontophoresis）的技术］。（B）采用短时但强烈的电击破坏细胞膜，使得细胞膜对那些物质有瞬时的通透性。（C）将要研究的物质装入封闭的膜微囊然后引导膜微囊和靶细胞融合。（D）DNA 包被的金颗粒用于将新基因导入细胞核

“笼状”前体分子的光感应活化作用促进了对细胞内动态的研究

许多细胞内过程具有复杂和快速的特性，譬如信号分子的活动或细胞骨架蛋白的运动，这使得在单细胞水平研究这些过程比较困难。理论上，人们希望能够将任何要研究的分子在精确的时间或位置导入到活细胞并能追踪它随后的行为以及细胞的应答。对投递位置和时间的控制的难度限制了显微注射。更有力的方法是合成要研究的分子的非活性态，将它导

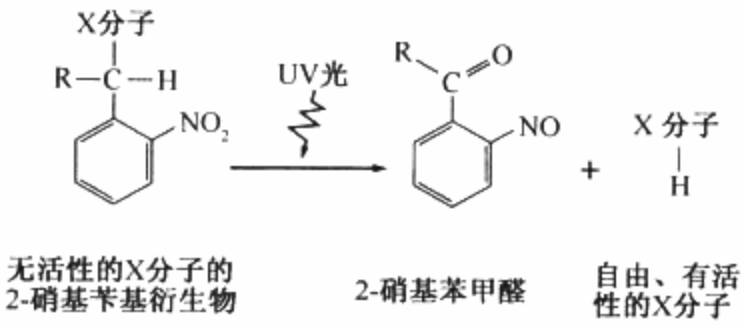


图 9-41　笼状分子

这幅图显示了一个分子（图中所指的 X）的光敏笼状衍生物如何被一紫外线转变成自由的活性形式。ATP 之类的小分子可以用这种方法封装。即使像 Ca^{2+} 之类的离子也能用非直接的方法封装；这种情况下使用能结合 Ca^{2+} 的螯合剂，螯合剂被光分解失活，释放出 Ca^{2+}

入细胞，然后通过聚焦一个光斑在细胞中选定的位点活化该分子。这种类型的非活性态的光敏前体叫做**笼状分子**，可以由包括 Ca^{2+}、cAMP、GTP 和三磷酸肌醇在内的多种的小分子构成。这些笼状分子可以通过图 9-40 中提到的方法导入细胞，然后通过一束激光脉冲活化（图9-41）。显微镜可用于将光脉冲聚焦到细胞任一微小区域，这样实验者能够正确控制分子投递的位置和时间。例如，用这种方法可以研究细胞内信号分子释放进入胞质的瞬时作用。

荧光分子也是笼状化时的重要工具。将光敏荧光染料结合到纯蛋白质上得到笼状分子。使用笼状荧光染料标记时蛋白质表面被加上了大量的大基团，这很容易改变蛋白质的性质，因此不同于放射性同位素标记（其中只有被标记的原子核内的中子发生变化），修饰后的蛋白质能保持生物活性非常重要。一个满意的标记方案通常需要反复试验才能建立。一旦得到了具有生物活性的标记蛋白质，它的行为可以在活细胞内被追踪。例如，用笼状荧光素标记的微管蛋白可以被并入有丝分裂纺锤体的微管中；当纺锤体的一小片区域被激光照亮时，标记的微管蛋白发出荧光，因此它沿纺锤体微管的运动可以被容易的追踪到（图 9-42）。从原理上看，相同的技术可以被用于任何蛋白质。

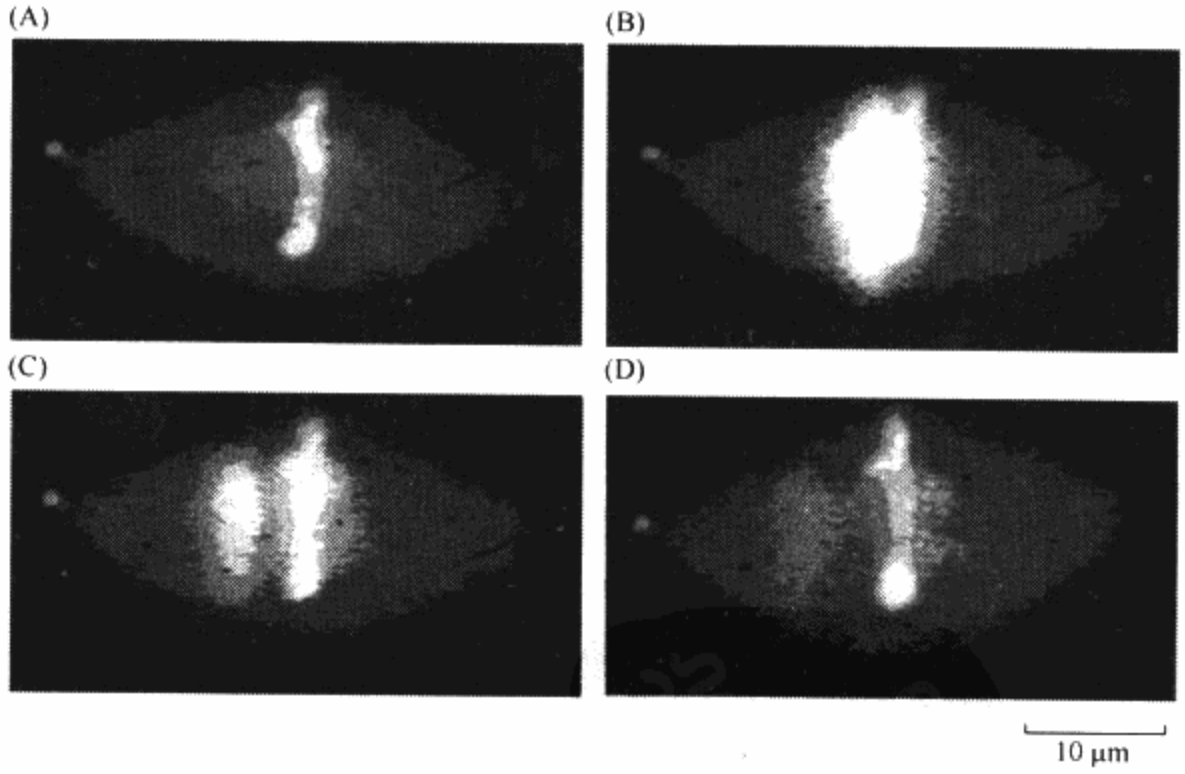


图 9-42　将笼形荧光素连接到微管蛋白上检测有丝分裂纺锤体中微管的变化

（A）从非洲蟾蜍（*Xenopus*）的卵中提取的有丝分裂中期的纺锤体上用 3 种荧光标记物结合：罗丹明标记的微管蛋白（红色）标记了所有的微管，一种蓝色的 DNA 绑定染料标记了染色体，笼形荧光素标记的微管蛋白也组合到所有的微管中，但是因为它是非荧光的，所以看不见，直到被紫外线激活。（B）一束紫外线用来在局部释放笼形荧光素标记微管蛋白，大部分刚好在赤道板的左侧。不久后［（C）中是 1.5min 后，（D）中是 2.5min 后］，释放的荧光素-微管蛋白信号被观察到朝左边的纺锤体移动，这显示了尽管纺锤体大体上没有变化（通过红色罗丹明标记的微管蛋白的荧光所观察到的），但微管蛋白仍持续向两极移动（图片摘自 K. E. Sawin and T. J. Mitchison, *J. Cell Biol.* 112：941～954，1991. © The Rockefeller University Press）

绿色荧光蛋白用于标记活细胞或组织中的蛋白质

上述讨论的所有荧光分子必须在细胞外制造，然后人工导入细胞。对那些自身固有荧光的蛋白质分子基因编码的发现，显示了大量新的机会。基因工程因而可以建立那些自带可视化标记的细胞系。这些细胞内指示分子通过发出荧光显示它们在胞内的运作。

细胞生物学中用于该目的荧光蛋白中最重要的是**绿色荧光蛋白（GFP）**，该蛋白质从水母 *Aequoria victoria* 中分离。这种蛋白质由一可克隆的单基因按常规方式编码，可以被导入其他种类的细胞中。刚翻译出的蛋白质并不是荧光的，但是大约在一小时内，经过自催化的翻译后修饰，该蛋白质形成了一保护在桶状蛋白质内部、高效的发荧光中心（图9-43）。对于最初的基因已经进行了大范围的位点定向突变，可以从动物、植物、真菌到微生物等生物体中获得荧光。荧光效率也被提高，得到了吸收和发射光谱在蓝绿黄范围内变化的变异体。最近在珊瑚虫中发现了荧光相关蛋白质的一个家族，因此将范围扩展到光谱的红光区域。

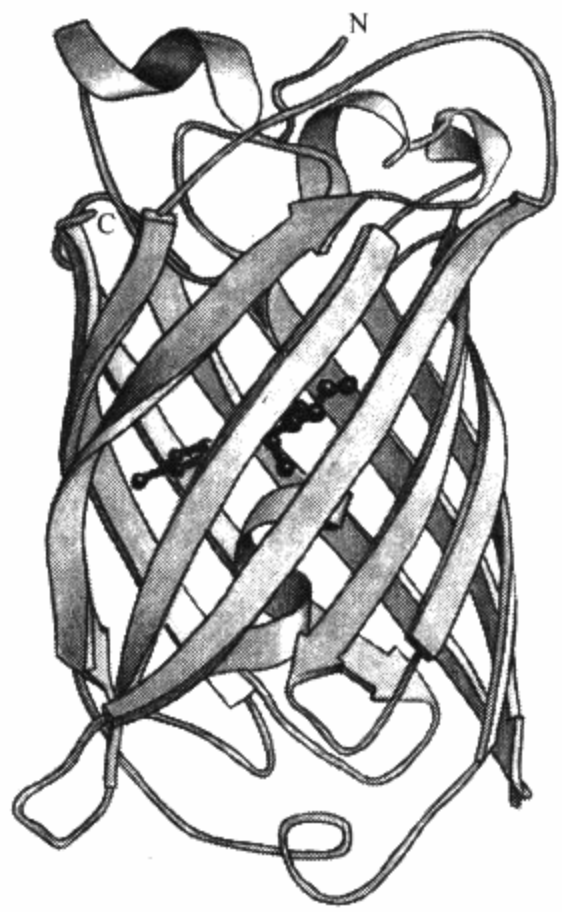


图 9-43　绿色荧光蛋白（GFP）

这里用示意图显示的 GFP 结构中最突出的部分是形成桶板的 11 条 β 链。藏在桶内部的是活性色素分子（暗绿色），这是通过 3 个氨基酸残基突出的侧链翻译后形成的（图片摘自 M. Ormö et al., *Science* 273：1392～1395，1995）

GFP 最简单的一个用途是作为监测基因表达的报道分子。要研究的基因的启动子用于控制 GFP 编码基因转录，构建一转基因生物；这将能在活体中直接给出基因表达模式的可见图像。另一应用是将肽定位信号加到 GFP 上，引导 GFP 到细胞特定部位，如内质网或线粒体，照亮这些细胞器使得可在生活状态进行观察。

GFP 的 DNA 编码序列也能被插入到另一蛋白质起始或终止的位置，得到由 GFP 域和那种蛋白质组合到一起的新蛋白质。在很多情形下 GFP 融合蛋白具有和原蛋白质相同的运转方式，可以直接显示原蛋白质的位置和活性（图 9-44）。通常可以证明 GFP 融合蛋白和未标记蛋白质在功能上是等效的。例如，可以用 GFP 融合蛋白来拯救原蛋白质缺失突变。在活生物体中 GFP 标记是显示蛋白质分布和动力学的最清晰明确的方法（图 9-45）。

GFP 及其类似物的用途正快速增加。例如，DNA 或 RNA 可以将结合某特定蛋白质的寡核苷酸序列并入它们的拷贝中，然后再同一细胞中表达加有 GFP 标签的结合蛋白。另一个用途是通过荧光共振能量转移（FRET）监测两种蛋白质的相互作用（图 8-49）。这项技术中，要研究的两个分子分别用不同的荧光染料标记，选择的两种荧光染料应该满足其中一种的发射光谱和另一种的吸收光谱交叠。如果这两种蛋白质结合使得它们所带荧光染料彼此能非常靠近（比大约 2nm 还近），吸收光的能量能够直接从一个荧光染料转移到另一个荧光染料。因此，当用第一个荧光染料激发波长的光照射复合物时，光会按第二个荧光染料的发射波长发出。使用具有不同光谱的 GFP 变种作为荧光染料，这种方法可以用于监测信号分子和它们

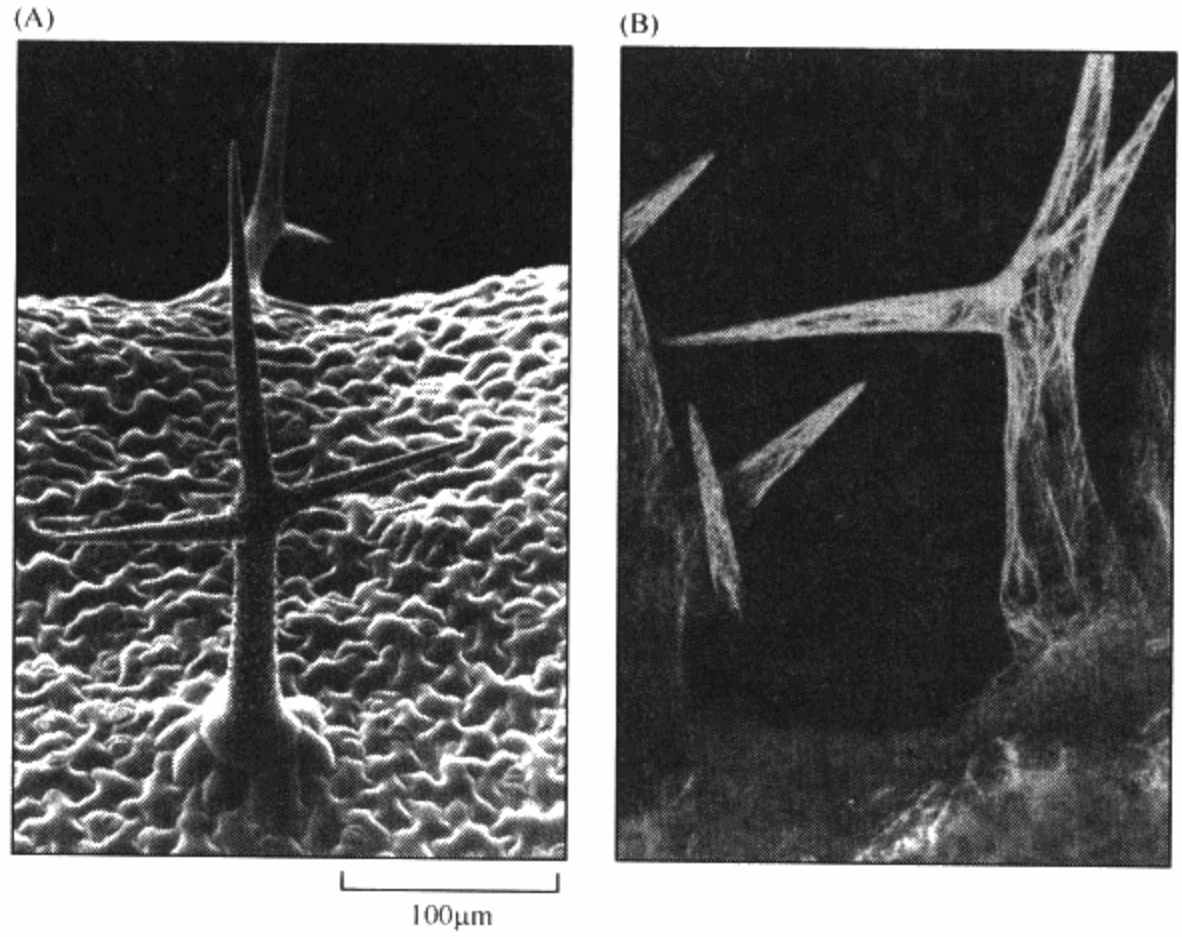


图 9-44　GFP 标签技术（GFP tagging）

（A）拟南芥叶子的上表面被大分岔的单细胞毛状物覆盖，这些毛状物是从表皮的表面起来的。这些毛状物或毛状体，能够通过 SEM 成像。（B）如果一个拟南芥中转入一个融合了 GFP 的 talin（一种和肌动蛋白结合的蛋白质）DNA 编码序列，在转基因植物的所有活细胞中产生的荧光 talin 蛋白会结合到肌动蛋白纤维上。共聚焦显微镜能够显示毛状体整个肌动蛋白细胞骨架的动态（绿色）。红色荧光是由叶表皮下的叶绿素产生的［（A）经 Paul Linstead 许可；（B）经 Jaideep Mathur 许可］

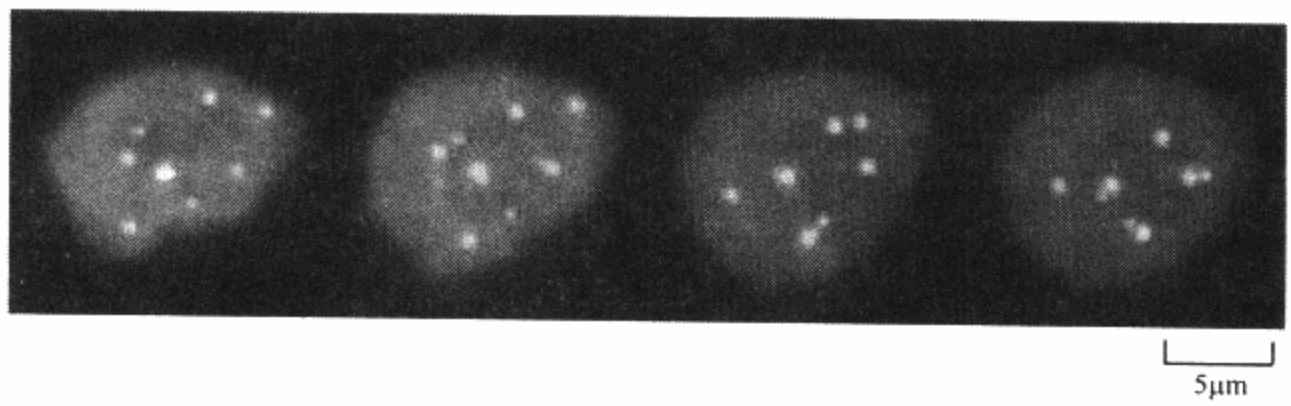


图 9-45　GFP 标记的动态

显微照片顺序显示了在一小时过程中拍摄的一活细胞核的一系列三维图像烟草细胞用融合了 GFP 的剪接体蛋白稳定转化，这种蛋白质被集中在叫做 Cajal 体的核小体中（图 6-48）。用激光共聚焦显微技术很容易看见荧光 Cajal 体在细胞核内四处运动的动态结构（经 Kurt Boudonck，Liam Dolan 和 Peter Shaw 许可）

受体之间的相互作用过程。

光除了用于对微小物体成像外还可用于操纵微小物体

光子带有少量动量。这意味着一个物体吸收或偏转一束光时会感受到微小的力。采用普通光源时，这种辐射压力太小没有意义。不过这在宇宙尺度很重要（有助于防止星体内部引

力塌缩），更普通点的是在细胞生物学实验室中，强烈聚焦的激光束能够施加足够大的力将周围微小物体推入细胞。如果激光束聚焦在一个折射率高于周围环境的某物体上时，光束被折射，导致大量光子改变方向。光子偏转方式将物体控制在光束的焦点；如果物体开始慢慢从这个位置散开，一边所受的辐射压比另一边高，物体被推回去。因此通过操作一聚焦激光束，通常用红外激光这种细胞成分吸收最低的激光，人们能建立“光学镊子”来移动细胞器或染色体等亚细胞物体。这种方法被用于测量由单个肌动蛋白-肌球蛋白分子、单个微管马达以及RNA聚合酶所施加的力。

能被生物材料更强吸收的强烈聚焦激光束也能被直接用作光子刀，用来杀死特殊的细胞、在细胞上切出或灼出孔或将细胞内的成分彼此分开。用这些方法，光学设备可作为用于细胞显微外科的基本工具盒。

可用放射性同位素标记的分子

正如我们所了解的，细胞生物学中测定特殊分子的量、了解特定分子在细胞中的位置以及对细胞外信号应答时分子含量和位置的变化，这些通常是很重要的。要研究的分子从 Ca^{2+} 或 H^+ 之类的无机离子到特定蛋白质、RNA 和 DNA 等大分子。到目前为止我们已经描述了用于分析这类分子的荧光方法的灵敏程度，也描述了如何跟踪活细胞中分子的动态行为。在本章的最后，我们描述放射性同位素如何用于显示特定分子通过细胞的路径。

大多数自然存在的元素是由轻微不同的同位素组成的混合物。它们的原子核的质量彼此有些差别，但是因为有着相同数量的质子和电子，所以它们的化学性质相同。放射性同位素中原子核是不稳定的，会经历随机裂变产生一不同的原子。在裂变的过程中，电子之类的高能亚原子颗粒或γ射线之类的放射线被发出。使用化学合成的方法将一个或多个放射性原子结合到要研究的糖或氨基酸等小分子中，在任何生物反应中那个分子可以被追踪。

尽管自然存在的放射性同位素很稀少（因为它们的不稳定性），但可以用高能粒子轰击稳定原子这一核反应大量生产放射性同位素。结果很容易得到很多生物学上重要元素的放射性同位素（表 9-3）。它们发出的射线可用多种方法监测。电子（β粒子）可以根据它们在气体中引起的电离用盖格计数器检测，或者用闪烁计数器检测在闪烁流体中它们引起的微小闪光。这种方法能够测量生物标本中存在的放射性同位素颗粒的精确数量。使用光学或电子显微镜，用放射自显影术也能检测标本中放射性同位素的位置。所有的这些检测方法都非常灵敏：在有利的环境中，几乎每次裂变，也就是说每个发生裂变的原子，都能被检测到。

表 9-3　生物学研究中常用的一些放射性同位素

同位素	半衰期
^{32}P	14 天
^{131}I	8.1 天
^{35}S	87 天
^{14}C	5570 天
^{45}Ca	164 天
^{3}H	12.3 天

注：同位素按照它们发出β射线（电子）的能量递减的顺序排列。^{131}I 也发出γ射线。半衰期是指 50%的同位素原子发生裂变所需要的时间。

放射性同位素作为示踪分子用于细胞和组织中

生物学中放射现象的最早应用之一是在光合作用中追踪碳的化学途径。单细胞绿藻养在包含有放射标记的 CO_2（$^{14}CO_2$）的大气中，将它们暴露在阳光下多次后，通过纸层析分离出它们的可溶性成分。小分子中包含有来自于 CO_2 的 ^{14}C，通过放置在干燥的纸色谱上的照相底片可检测到这些。用这种方法可以鉴定从 CO_2 到糖类等光合作用途径中主要的成分。

放射性分子几乎可以用于追踪细胞中任一活动的过程。在一典型实验中将放射性前体分子供应给细胞。放射性分子和先前未被标记的分子混合在一起；因为它们仅仅在原子核的重量上有轻微差别，所以细胞处理这些分子是一致的。放射性分子的位置和化学形式的变化可以实时监测。使用**脉冲追踪**标记方案能够提高这种实验的分辨率，这种方法中放射性物质（脉冲剂）加入非常短的时间后被洗掉，然后重新加入非放射性的分子（追踪剂）。按规则的时间间隔取样，可鉴定每个样品中放射性的化学形态或位置（图 9-46）。例如，将脉冲追踪实验和放射自显影相结合，对于阐明蛋白质从内质网到细胞外部的分泌途径有重要意义(图 9-47)。

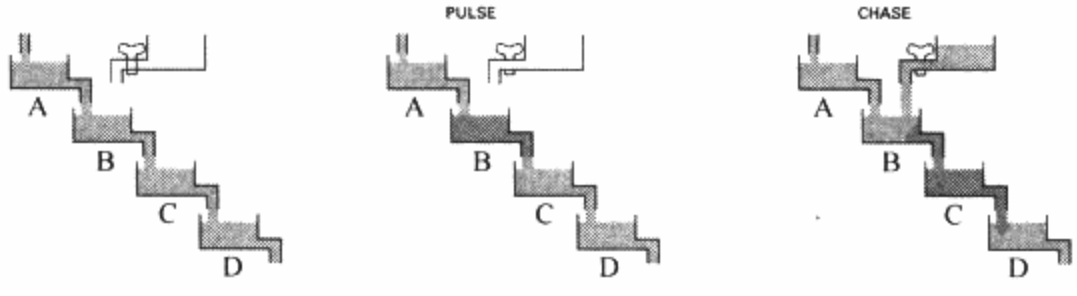


图 9-46　一个使用放射性同位素脉冲-追踪的典型试验

标记为 A、B、C 和 D 的 4 个腔表示细胞中不同的部分（通过放射自显影或细胞分离实验测定）或不同的化学成分（通过色谱或其他化学方法测定）。一个真实的脉冲-追踪实验的结果见图 9-47

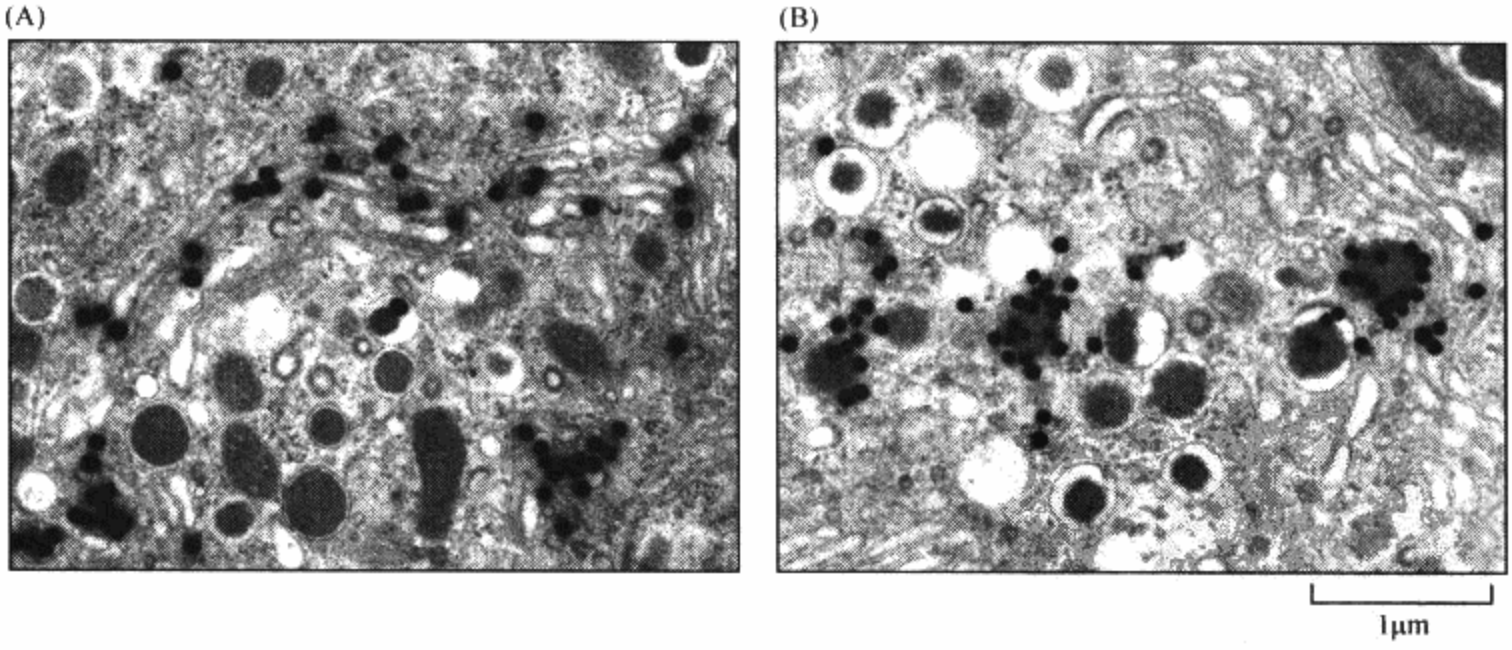


图 9-47　电子显微镜放射自显影成像术

在胰腺 B 细胞的脉冲-追踪实验中，首先用 ^3H-亮氨酸孵育 5min（脉冲），接着使用过量未标记的亮氨酸（追踪）。该氨基酸主要用于合成胰岛素，然后被分泌。追踪实验发现，10min 后，标记蛋白质已经从粗糙内质网移动到高尔基体（A），在照相乳液中形成的黑色银颗粒显示蛋白质所在的位置。45min 后，标记蛋白质发现于电子致密的分泌颗粒中（B）。图中显示的圆形微小银颗粒由一种特殊的显影剂产生，不要与免疫金标记技术中看起来类似的黑色小点混淆（图 9-26）。与此类似的实验在建立新合成的分泌蛋白所采取的胞内途径中是重要的（经 L. Orci 许可，摘自 *Diabetes* 31：538～565，1982. © American Diabetes Association, Inc）

放射性同位素标记方法在区别化学组成相同但经历不同（如在合成时间上不一样）的分子方面具有独特的价值。通过这种方法，可以显示活细胞中几乎所有的分子都不断地被降解和替换，即使细胞已经不再生长明显处于稳态了。这种“周转”有时发生的非常慢，没有放射性同位素几乎不可能检测出来。

现在几乎所有的常见小分子的放射性形式都能从商业来源获得，实际上无论多复杂的生物分子都能用放射性标记。可以将放射性原子并入化合物结构的特定部位，使得能够追踪生物反应中同一分子不同部分分离的情况(图 9-48)。

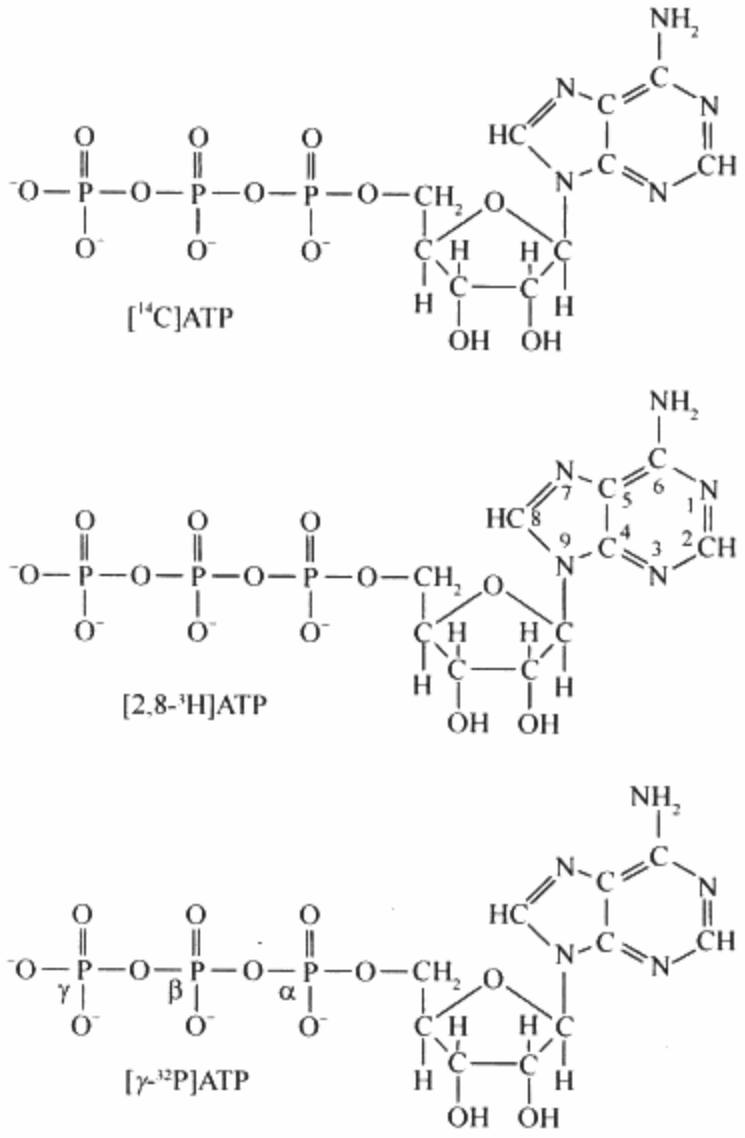


图 9-48　放射性同位素标记的分子

商业上提供的 ATP 的 3 种放射类型，其中红色显示的是放射性原子。图中也显示了鉴定放射性原子的位置和类型的术语

正如前面提到的，放射现象在细胞生物学中的一个重要用途是使用**放射自显影法**使放射性化合物局限于整个细胞或组织的局部。在这个过程中，按光学或电子显微成像方法进行固定和处理之前，活细胞被暂时地暴露在一特定放射性化合物的脉冲下，培养适当的时间，使得它们能够并入那些化合物。然后当放射性同位素衰变时使用照相的摄影乳剂薄胶片覆盖每个标本，搁置在暗处几天。然后将摄影乳剂显影，显影的银颗粒的位置显示了每个细胞中放射性的位置（图 5-33）。如果细胞被暴露在^{3}H-胸腺嘧啶脱氧核糖核苷（DNA 的一个放射性前体）中，可以显示 DNA 在细胞核中的合成和并保持在那里。相反，如果细胞被暴露在^{3}H-尿嘧啶核苷（RNA 的一个放射性前体）中，会发现 RNA 现在细胞核内合成然后被快速转移到细胞质中。通过凝胶电泳将分子分开后，放射性标记的分子也能通过放射自显影方法检测：使用这种方法蛋白质（图 8-17）和核酸（图 8-23）在凝胶上的位置通常都能被检测出来。

小结

利用现有技术几乎可以探测、测量和追踪活细胞中任一想研究的分子。例如，荧光指示剂染料能够用于测量单个细胞或细胞中不同部分的特定离子的浓度。通过构建无活性的“笼状”前体可以追踪活细胞中许多分子的动态行为，那些前体可以被导入细胞，然后在细胞中某选定区域被光刺激反应激活。绿色荧光蛋白（GFP）通过基因操作能附着其他蛋白质，是特别通用的探针。实际上任何要研究的蛋白质都能通过基因工程得到 GFP 融合蛋白，然后通过荧光显微成像术在活细胞中成像。多种元素的放射性同位素也被用于生物化学和显微镜学中追踪特定分子的命运。

参考文献

一般文献

Celis JE (ed) (1998) Cell Biology: A Laboratory Handbook, 2nd edn. San Diego: Academic Press.

Lacey AJ (ed) (1999) Light Microscopy in Biology: A Practical Approach, 2nd edn. Oxford: Oxford University Press.

Paddock SW (ed) (1999) Methods in Molecular Biology, vol 122: Confocal Microscopy Methods and Protocols. Totowa, NJ: Humana Press.

Watt IM (1997) The Principles and Practice of Electron Microscopy, 2nd edn, Cambridge: Cambridge University Press.

Looking at the Structure of Cells in the Microscope

Agard DA & Sedat JW (1983) Three-dimensional architecture of a polytene nucleus. *Nature* 302, 676–681.

Agard DA, Hiraoka Y, Shaw P & Sedat JW (1989) Fluorescence microscopy in three dimensions. In Methods in Cell Biology, vol 30: Fluorescence Microscopy of Living Cells in Culture, part B (DL Taylor, Y-L Wang eds). San Diego: Academic Press.

Allen RD (1985) New observations on cell architecture and dynamics by video-enhanced contrast optical microscopy. *Annu. Rev. Biophys. Chem.* 14, 265–290.

Allen TD & Goldberg MW (1993) High resolution SEM in cell biology. *Trends Cell Biol.* 3, 203–208.

Boon ME & Driver JS (1986) Routine Cytological Staining Methods. London: Macmillan.

Böttcher B, Wynne SA & Crowther RA (1997) Determination of the fold of the core protein of hepatits B virus by electron cryomicroscopy. *Nature* 386, 88–91.

Celis JE (ed) (1998) Cell Biology: A Laboratory Handbook, 2nd edn, vol 3, part 1—Histology and Histochemistry, pp 219–248. San Diego: Academic Press.

Celis JE (ed) (1998) Cell Biology: A Laboratory Handbook, 2nd edn, vol 2, part 6—Immunocytochemistry, pp 457–494. San Diego: Academic Press.

Celis JE (ed) (1998) Cell Biology: A Laboratory Handbook, 2nd edn, vol 3, part 9—Electron Microscopy, pp 249–362. San Diego: Academic Press.

Dubochet J, Adrian M, Chang J-J et al. (1988) Cryoelectron microscopy of vitrified specimens. *Q. Rev. Biophys.* 21, 129–228.

Everhart TE & Hayes TL (1972) The scanning electron microscope. *Sci. Am.* 226, 54–69.

Fowler WE & Aebi U (1983) Preparation of single molecules and supramolecular complexes for high resolution metal shadowing. *J. Ultrastruct. Res.* 83, 319–334.

Gabashvili IS, Agrawal RK, Spahn CMT et al. (2000) Solution structure of the *E. coli* 70S ribosome at 11.5 Å resolution. *Cell* 100, 537–549.

Griffiths G (1993) Fine Structure Immunocytochemistry. Heidelberg: Springer-Verlag.

Harlow E & Lane D (1988) Antibodies. A Laboratory Manual. Cold Spring Harbor: Cold Spring Harbor Laboratory.

Haugland RP (ed) (1996) Handbook of Fluorescent Probes and Research Chemicals, 8th edn. Eugene, OR: Molecular Probes, Inc. (Available online at http://www.probes.com/)

Hayat MA (1978) Introduction to Biological Scanning Electron Microscopy. Baltimore: University Park Press.

Hayat MA (1981) Fixation for Electron Microscopy. New York: Academic Press.

Hayat MA (1989) Cytochemical Methods. New York: Wiley-Liss.

Hayat MA (2000) Principles and Techniques of Electron Microscopy, 4th edn. Cambridge: Cambridge University Press.

Hecht E (2002) Optics, 4th edn. Reading, MA: Addison-Wesley.

Heitlinger E, Peter M, Haner M et al. (1991) Expression of chicken lamin B2 in *Escherichia coli*: characterization of its structure, assembly and molecular interactions. *J. Cell Biol.* 113, 485–495.

Heuser J (1981) Quick-freeze, deep-etch preparation of samples for 3-D electron microscopy. *Trends Biochem. Sci.* 6, 64–68.

Hoenger A & Aebi U (1996) Three-dimensional reconstructions from ice-embedded and negatively stained macromolecular assemblies: a critical comparison. *J. Struct. Biol.* 117, 99–116.

Inoué S & Spring HR (1997) Video Microscopy, 2nd edn. New York: Plenum Press.

Kuhlbrandt W, Wang DN & Fujiyoshi Y (1994) Atomic model of plant light-harvesting complex by electron crystallography. *Nature* 367, 614–621.

Mancini EJ, Clarke M, Gowen BE, Rutten T & Fuller SD (2000) Cryo-electron microscopy reveals the functional organization of an enveloped virus, Semliki Forest virus. *Mol. Cell* 5, 255–266.

Maunsbach AB (1998) Immunolabelling and staining of ultrathin sections in biological electron microscopy. In Cell Biology: A Laboratory Handbook, vol 3 (JE Celis ed), pp 268–276. San Diego: Academic Press.

Minsky M (1988) Memoir on inventing the confocal scanning microscope. *Scanning* 10, 128–138.

Pawley JB (ed) (1995) Handbook of Biological Confocal Microscopy, 2nd edn. New York: Plenum Press.

Pease DC & Porter KR (1981) Electron microscopy and ultramicrotomy. *J. Cell Biol.* 91, 287s–292s.

Petran M, Hadravsky M, Egger MD & Galambos R (1968) Tandem-scanning reflected light microscope. *J. Opt. Soc. Am.* 58, 661–664.

Ploem JS & Tanke HJ (1987) Introduction to Fluorescence Microscopy. Royal Microscopical Society Microscopy Handbook, No 10. Oxford: Oxford University Press.

Shotton DM (1998) Freeze fracture and freeze etching. In Cell Biology: A Laboratory Handbook, vol 3 (JE Celis ed), pp 310–322. San Diego: Academic Press.

Shotton DM (ed) (1990) Electronic Light Microscopy: Techniques in Modern Biomedical Microscopy. New York: Wiley-Liss.

Slayter EM & Slayter HS (1992) Light and Electron Microscopy. Cambridge: Cambridge University Press.

Taylor DL & Wang Y-L (eds) (1989) Methods in Cell Biology, vols 29 & 30: Fluorescence Microscopy of Living Cells in Culture, parts A & B. San Diego: Academic Press.

Unwin PNT & Henderson R (1975) Molecular structure determination by electron microscopy of unstained crystal specimens. *J. Mol. Biol.* 94, 425–440.

Weiss DG, Maile W, Wick RA & Steffen W (1999) Video microscopy. In Light Microscopy in Biology: A Practical Approach, 2nd edn (AJ Lacey ed). Oxford: Oxford University Press.

White JG, Amos WB & Fordham M (1987) An evaluation of confocal versus conventional imaging of biological structures by fluorescence light microscopy. *J. Cell Biol.* 105, 41–48.

Wigge PA, Jensen ON, Holmes S et al. (1998) Analysis of the Saccharomyces spindle pole by matrix-assisted laser desorption/ionization (MALDI) mass spectrometry. *J. Cell Biol.* 141, 967–977.

Wilson L & Matsudaira P (eds) (1993) Methods in Cell Biology, vol 37: Antibodies in Cell Biology. New York: Academic Press.

Zernike F (1955) How I discovered phase contrast. *Science* 121, 345–349.

Visualizing Molecules in Living Cells

Ammann D (1986) Ion-Selective Microelectrodes: Principles, Design and Application. Berlin: Springer-Verlag.

Celis JE (ed) (1998) Cell Biology: A Laboratory Handbook, 2nd edn, vol 4: part 13—Transfer of Macromolecules and Small Molecules, pp 3–183. San Diego: Academic Press. (This contains many descriptions of the variety of methods that are used in various different cell types for different purposes, including microinjection, electroporation, liposome delivery, biolistic bombardment and other methods.)

Celis JE (ed) (1998) Cell Biology: A Laboratory Handbook, 2nd edn, vol 3, part 10—Intracellular Measurements, pp 363–386. San Diego: Academic Press.

Chalfie M, Tu Y, Euskirchen G, Ward WW & Prasher DC (1994) Green fluorescent protein as a marker for gene expression. *Science* 263, 802–805.

Haseloff J, Dormand E-L & Brand A (1999) Live imaging with green fluorescent protein. In Methods in Molecular Biology, vol 122: Confocal Microscopy Methods and Protocols (SW Paddock ed). Totowa, NJ: Humana Press.

Mitchison TJ & Salmon ED (1992) Kinetochore fiber movement contributes to anaphase-A in newt lung cells. *J. Cell Biol.* 119, 569–582.

Neher E (1992) Ion channels for communication between and within cells. *Science* 256, 498–502.

Ormo M, Cubitt AB, Kallio et al. (1966) Crystal structure of the *Aequoria*

victoria green fluorescent protein. *Science* 273, 1392–1395.

Parton RM & Read ND (1999) Calcium and pH imaging in living cells. In Light Microscopy in Biology. A Practical Approach, 2nd edn (Lacey AJ ed). Oxford: Oxford University Press.

Sakmann B (1992) Elementary steps in synaptic transmission revealed by currents through single ion channels. *Science* 256, 503–512.

Sawin KE, Theriot JA & Mitchison TJ (1999) Photoactivation of fluorescence as a probe for cytoskeletal dynamics in mitosis and cell motility. In Fluorescent and Luminescent Probes for Biological Activity, 2nd edn (Mason WT ed). San Diego: Academic Press.

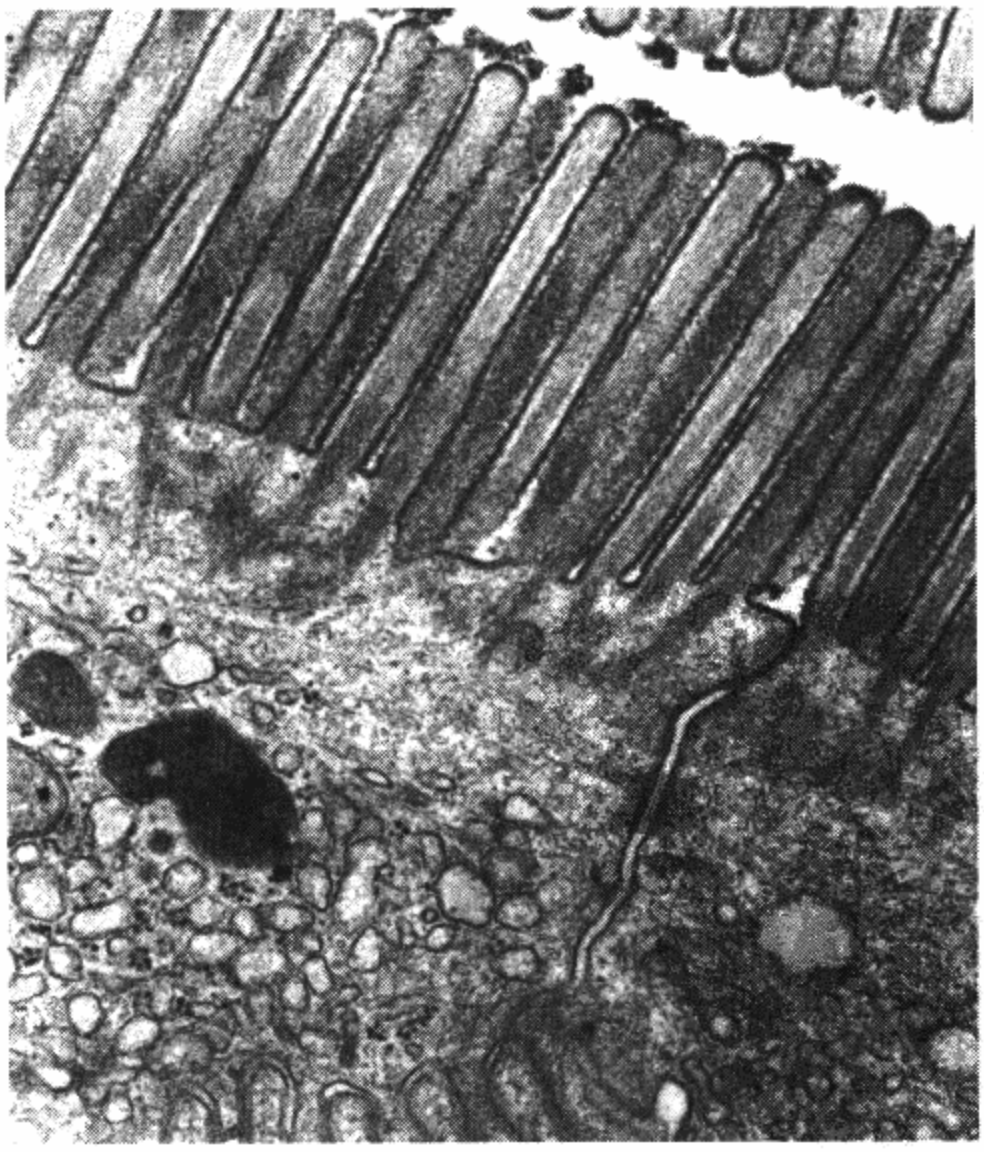
Sheetz MP (ed) (1997) Methods in Cell Biology, vol 55: Laser Tweezers in Cell Biology. San Diego: Academic Press.

Theriot JA, Mitchison TJ, Tilney LG & Portnoy DA (1992) The rate of actin-based motility of intracellular *Listeria monocytogenes* equals the rate of actin polymerization. *Nature* 357, 257–260.

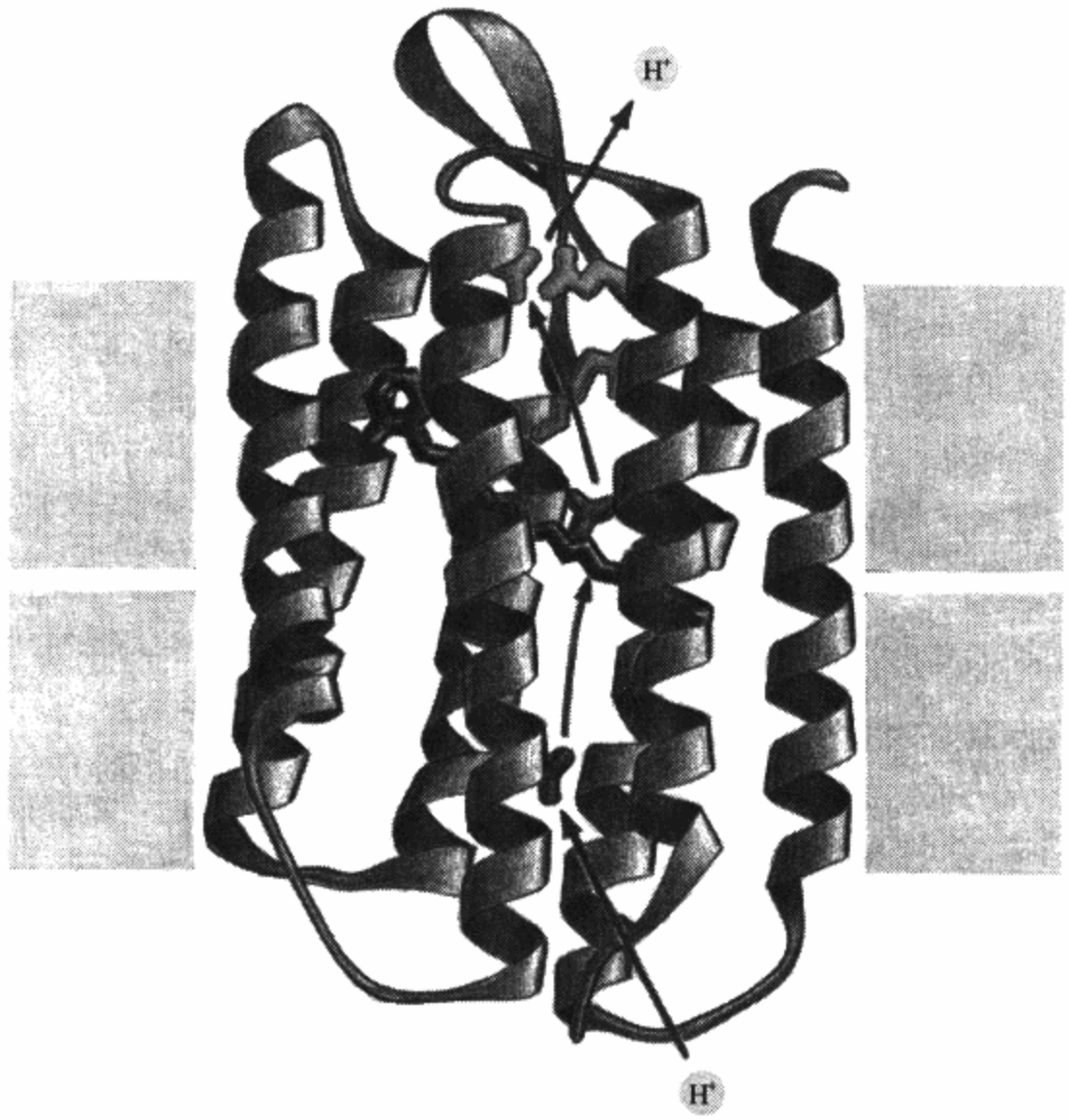
Tsien RY (1989) Fluorescent probes of cell signaling. *Ann. Rev. Neurosci.* 12, 227–253.

第Ⅳ部分　细胞的内部构造

10　膜结构
11　小分子的膜运输和膜的电性质
12　细胞内区域和蛋白质分选
13　细胞内的膜泡运输
14　能量转换：线粒体和叶绿体
15　细胞通讯
16　细胞骨架
17　细胞周期与程序性细胞死亡
18　细胞分裂的机制



细胞质的组织



膜蛋白。插入到细胞膜中的特殊蛋白产生的空，可以使分子顺利通过。

10 膜 结 构

脂双层
膜蛋白

脂双层

细胞膜对于细胞的生存是至关紧要的。**质膜**环绕着细胞，确定了细胞的轮廓，保持细胞质和细胞外环境的本质不同。在真核细胞中，膜能够维持内质网、高尔基体、线粒体以及其他有膜包被的细胞器中的内含物和细胞质之间的特性差异。膜上有专门的蛋白质能形成跨膜离子梯度，这种梯度通常能帮助合成 ATP，驱使某些溶质的跨膜运动或者在神经及肌肉细胞中产生并传导电信号。在所有细胞中，质膜上也有些蛋白质作为外部信号的感受器，使细胞通过改变自身行为对环境信号做出应答；这些蛋白质感受器（或者称做受体）跨膜传递的信号不仅仅是离子或分子。

无论生物膜的功能多么不同，所有的膜结构大致一样：都是由一非常薄的脂质膜和蛋白质分子组成，蛋白质分子和脂质膜主要是通过非共价相互作用结合在一起。细胞膜是流动的结构，大多数上面的分子能在膜平面上移动。脂类分子排列成 5nm 厚的连续双层（图 10-1）。脂双层构成了膜流动结构的基础，作为不透水层阻挡了大多数水溶性分子通过。嵌在脂双层中的跨膜蛋白质分子几乎行使了膜的其他所有功能，例如，有些蛋白质分子运输特定分子穿过膜，有些能催化合成 ATP 等膜相关反应。质膜中，有些蛋白质穿过脂双层将细胞骨架和细胞外基质或邻近的细胞连接起来，而某些其他蛋白质作为受体可检测并转换细胞周围环境中的化学信号。正如人们预料的，有许多不同的膜蛋白使细胞能行使功能并和环境发生相互作用。实际上，由动物细胞基因组编码的蛋白质中，估计大约有 30%的是膜蛋白。

图 10-1 细胞膜的三幅图

（A）一个质膜（人红细胞的）截面的电子显微镜照片，（B）和（C）这些示意图显示了细胞膜二维和三维的图像（A，经 Daniel S. Friend 许可）

本章我们主要涉及生物膜的两种主要成分（脂质和膜蛋白）的结构和构成。尽管我们主要集中在细胞质膜方面，但涉及的大多数概念对于细胞中的各种各样的内在的膜也是适用的。细胞膜的功能在后续的章节中讨论。例如，第 14 章中将讨论细胞膜担任 ATP 合成的角色；第 11 章中将讨论细胞膜担负小分子跨膜运输的任务；第 15 章和第 19 章中将分别讨论细胞膜担负的细胞信号和细胞粘连任务。在第 12 章和第 13 章中，我们要讨论细胞内部的膜以及蛋白质的通过或嵌入膜。

膜脂类分子是两性分子，大多数能自发形成双分子层

脂类分子占大多数动物细胞膜质量的 50%，剩余的大部分是蛋白质。在 $1\mu m \times 1\mu m$ 面积的脂双层中大约有 5×10^6 个脂类分子，在一个小的动物细胞质膜中大约有 10^9 个脂类分子。细胞膜中所有的脂类分子是**两性**分子，也就是说这些分子具有一个**亲水**的极性端和一个**疏水**的非极性端。

含量最丰富的膜脂类分子是**磷脂**。这些分子具有一个极性的头部基团和两个疏水的羟基尾。这尾巴通常是脂肪酸，长度上可以不同（通常包含的碳原子数为 14～24 个）。一个尾巴通常具有一个或更多个顺式双键（也就是不饱和的），不过另外一个尾巴没有双键（也就是饱和的）。正如图 10-2 中显示的，每个双键在尾巴中造成一个小纽结。脂肪酸尾在长度和饱和度上的差异是重要的，因为影响了磷脂分子彼此之间结合的能力，因而影响了膜的流动性。

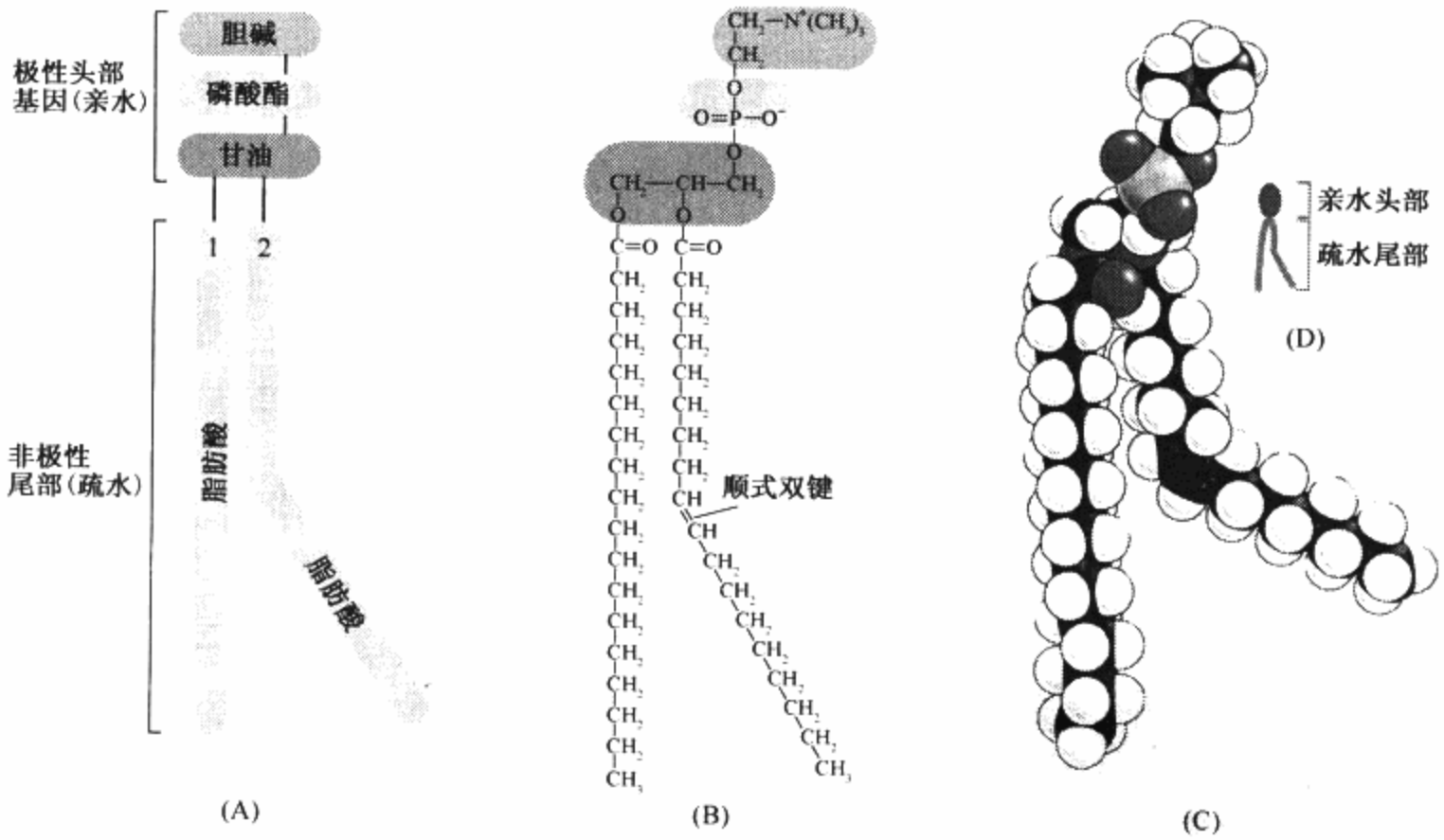


图 10-2　一个磷脂分子的各部分

这里给出的是卵磷脂，分别用以下来表示：（A）示意图；（B）结构式；（C）分子空间结构模型；（D）符号。为了突出效果，图中夸大了由顺式双键产生的扭结

由于脂分子的形状和两性分子的性质使得它们在水相环境中能够自发的形成双分子层。正如第 2 章所讨论的，亲水分子能够容易地溶解在水里，是因为它们含有带电基团或不带电的极性基团，这些基团能够和水分子形成静电相互作用或者氢键。相反，疏水分子在水中难溶是因为这些分子几乎所有的原子都是不带电的或者非极性的，因此不能和水分子形成较好的静电相互作用。如果分散在水中，它们会推动周围的水分子重新组织成环绕在这些疏水分

子周围的冰状笼（图 10-3）。因为这些笼状结构比周围的水更有序，它们的形成提高了自由能。然而，如果疏水分子（或者亲水分子的疏水部分）能成簇排列在一起，以至于最少数目的水分子受到影响，这能使得自由能的提高最小化。

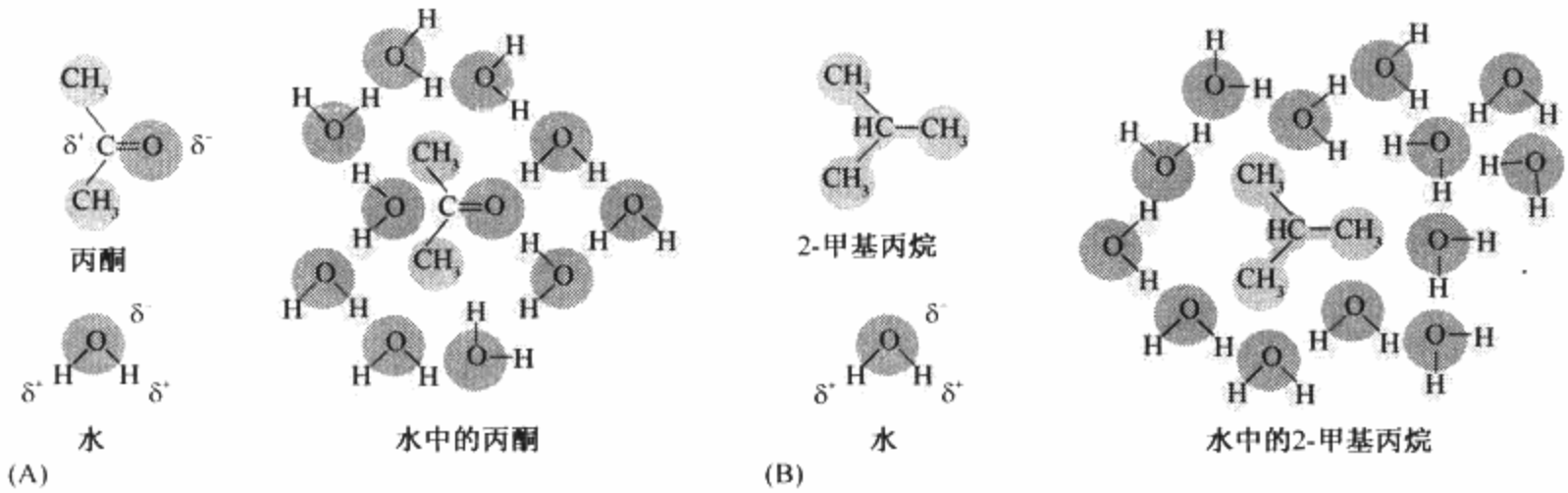


图 10-3　亲水分子和疏水分子与水的作用有何不同

（A）因为丙酮有极性，它能和同样具有极性的水分子形成良好的静电相互作用，因此丙酮很容易溶于水。（B）相反，2-甲基丙烷是完全疏水的。它不能和水形成良好的相互作用，它将邻近的水分子驱赶到一起重组成冰状的笼形结构，这提高了自由能。因此这种成分在水中实质上是不溶的。δ^- 表示一局部负电，δ^+ 表示一局部正电。极性原子用彩色显示，非极性基团用灰色显示

因为上述原因，脂类分子自发聚集将疏水尾部埋在内部，极性头部暴露在水中。根据这些分子的形状，这些分子能够通过以下两种方法之一来实现这一点：脂类分子能形成球形胶囊，尾巴在内部；或者脂类分子能形成双层膜，形成疏水性尾巴夹在亲水性头部基团之间的三明治结构（图 10-4）。

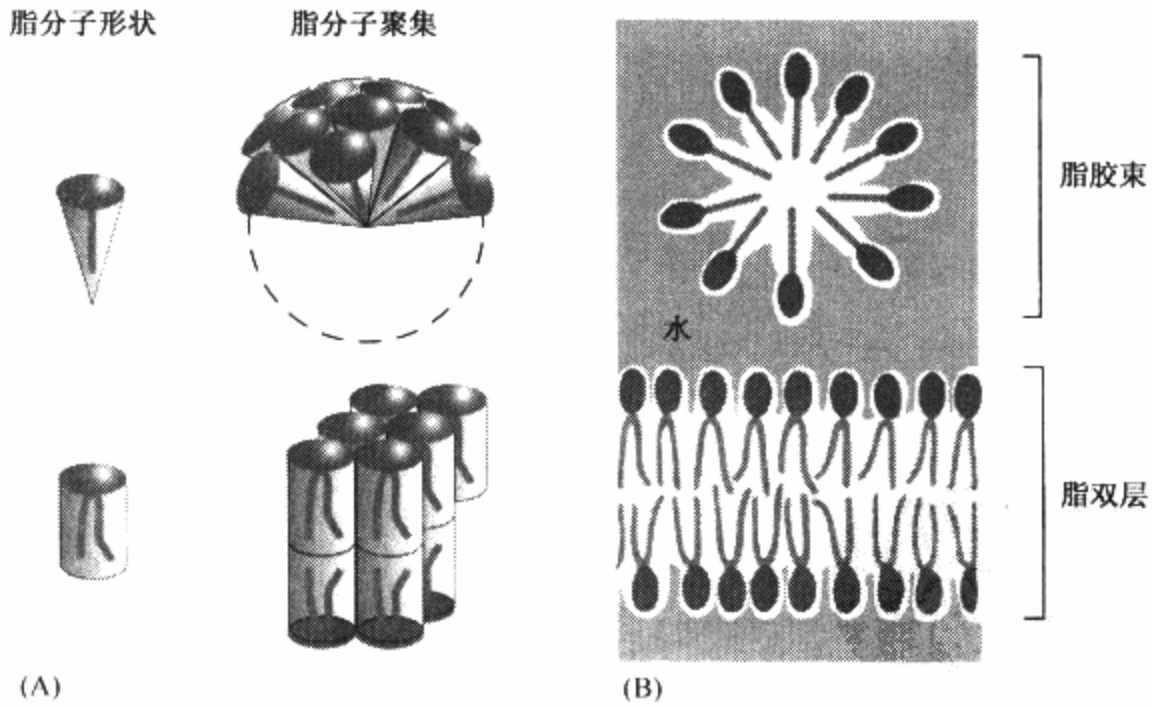


图 10-4　在水环境中脂分子的紧密排列

（A）楔形脂分子（上图）形成胶束，而圆柱形磷脂分子（下图）形成双分子层。（B）脂胶束和脂双层的横截面观。根据它们的形状，脂分子在水中自发形成一个或更多这类结构

因为是圆柱形的，磷脂分子在水环境中自发地形成双分子层。在这种能量上最有利的排列中，在双分子层的每一面都是亲水头部朝向水中，疏水尾部藏在中间和水隔开。这种驱使磷脂分子形成双分子层的力也提供了一种自愈的性质。在双分子层中的一个小裂口造成了一

个由水形成的自由边；因为这在能量上是不利的，脂分子自发重排消除自由边。（在真核细胞质膜中，大的裂口由细胞内的囊泡修复。）这种自由边的终结有重要的地位：双分子层避免这种缺口的唯一方法是自身闭合或者形成密封腔（图 10-5）。这种非凡的行为是活细胞产生的基础，依靠的是磷脂分子形状和两性分子的特性。

图 10-5　脂双层自发闭合形成一个密封腔

这种封闭结构是稳定的，因为它避免了疏水烃链暴露到水中，这种暴露在能量上是不利的

脂双层的自密封性质使得它成为细胞膜的理想结构，脂双层除了这一性质外还有其他特性。这些特性中最重要的一个就是流动性，这对于膜的很多功能是至关重要的。

脂双层是二维流体

大概在 1970 年左右，研究者第一次证实了单个脂分子能在脂双层中自由扩散。最初的证明来自于对合成类脂双层的研究。在这样的研究中两类制剂非常有用：①由球形囊形状形成的双分子层，叫做**脂质体**，根据制作方式的不同可以获得直径从 25nm～1μm 大小的多种球形囊（图 10-6）；②平面双分子层，叫做**黑膜**（black membrane），横跨两个水层之间隔膜上的小洞排列而成（图 10-7）。

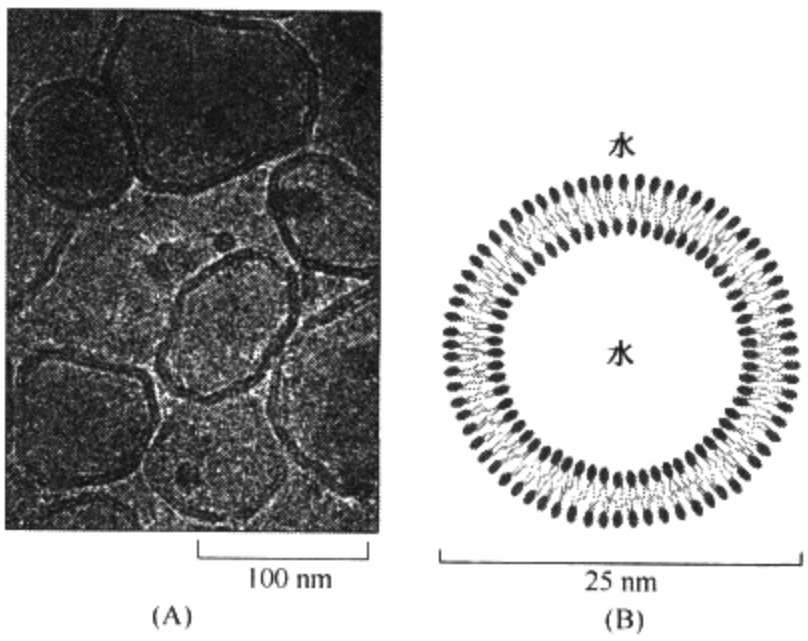


图 10-6　脂质体

（A）水中的一个未固定、未染色的磷脂泡囊在液氮温度下被快速冷冻。脂质体双分子层结构容易被看见到。（B）一个小球形脂质体横截面的示意图。在实验性研究中脂质体通常被用作典型膜［（A）经 Jean Lepault 许可］

多种技术已经被用于测量单个脂分子以及它们不同部分的运动。例如，其中一种技术是构造一个脂质分子，该分子的极性头部基团带有一个“自旋标记”，如硝酰基（$\diagdown\!\!\diagup$N—O$\diagdown\!\!\diagup$）；

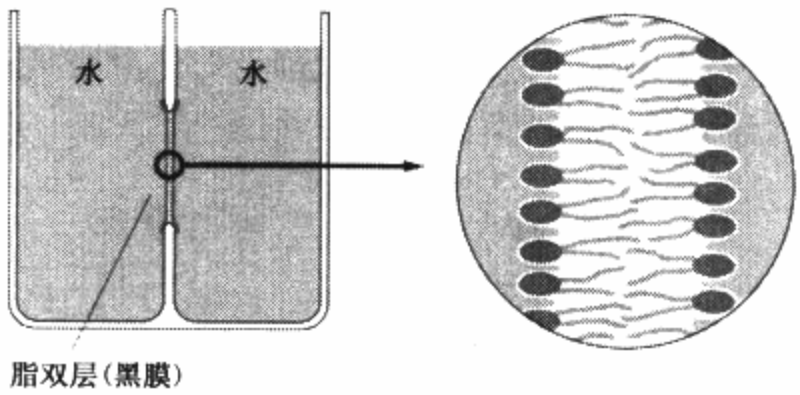


图 10-7　一个黑膜（合成的脂双层）的横截面图像

在两个分离的水区室之间的孔上形成的平面双分子层看上去是黑色的。黑膜被用于检测合成膜的渗透性

这包含了一个不配对的电子，该电子的自旋引起了一个超顺磁性的信号，这个信号可以基于电子自旋共振谱法（ESR）检测到（这种技术的原理和第 8 章讨论的核磁共振的原理类似）。一个自旋标记的脂分子在脂双层中的运动和方向可以通过 ESR 光谱推断出来。这类研究显示了在合成的双层中的磷脂分子很少能从一边的单层（也叫做 leaflet）移动到另一边的单层。对于任何单独的脂分子，这种翻转的过程出现的概率少于一个月一次。相对的，在膜的单层中脂分子和它们周围的邻居快速交换着位置（约每秒 10^7 次）。这引起了快速的径向扩散，扩散系数（D）大约是 10^{-8} cm^2/s，这意味着一个通常的脂分子能在一秒钟之内扩散一个大的细菌细胞的长度（2μm 左右）。这些研究也显示了单个的脂分子在它们的长轴附近快速旋转，以及它们的烃链具有很好的柔韧性（图 10-8）。

在单个的生物膜和活细胞中也采用标记后的脂分子完成了类似的研究。结果通常和在合成的双分子层中的相同，这些结果证明了生物膜的脂质成分是二维流体，其中组成分子可以自由的侧向移动。由于在合成的双分子层中，单独的磷脂分子通常被限制在自身所属的单分子层上。这一限制对于它们的合成产生了一个问题。磷脂分子只在膜的一个单分子层中产生，主要是在内质网（ER）膜的细胞质这一面的单分子层中产生。如果这些新产生的分子不能相当快速地迁移到非细胞质那一面的单分子层中，新的脂双层不可能产生。这一问题由一类特殊的结合在膜上的酶——磷脂转运蛋白来解决，这些酶催化了磷脂分子从一个单分子层到另一单分子层的翻转，第 12 章还要具体描述这种酶。

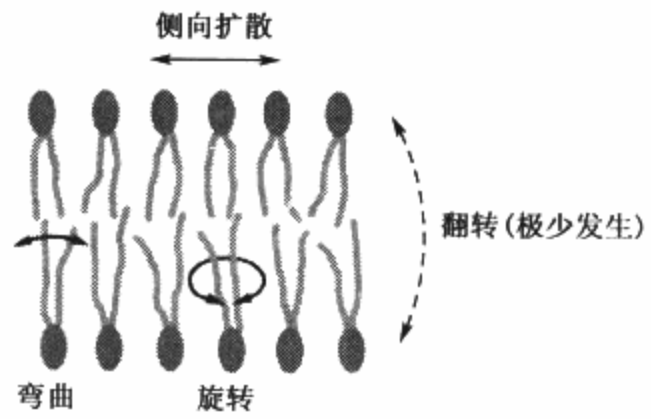


图 10-8　磷脂分子的活动性

在脂双层中磷脂分子可能的运动类型

脂双层的流动性依赖于它的成分

细胞膜的流动性必须得到精确地调控。例如，当双分子层的黏性在实验中提高到临界水平之上时，某些膜转运过程和酶的活性就停止了。

脂双层的流动性依赖于它的成分和温度，这在合成的双分子层的研究中很容易得到证明。在一个特定的凝固点，一个由单一类型磷脂分子构成的合成双分子层会从液态变成二维刚性的结晶（或凝胶）状态。这种状态变化叫做相变，如果烃链更短或有双键存在，这个相变温度就会更低（也就是说膜变得更难冷冻）。更短的烃链长度降低了烃链尾巴彼此相一致的趋势，顺式双键在烃链中产生的扭结使得这些分子更难聚在一起，所以膜能在更低的温度还保

持流动性（图 10-9）。细菌、酵母和其他一些温度随环境变化的生物体通过调整膜脂质中脂肪酸的成分来保持相对持续的流动性。例如，当温度降低时，具有更多顺式双键的脂肪酸被合成，这样就可以避免温度降低所导致的双分子层流动性的降低。

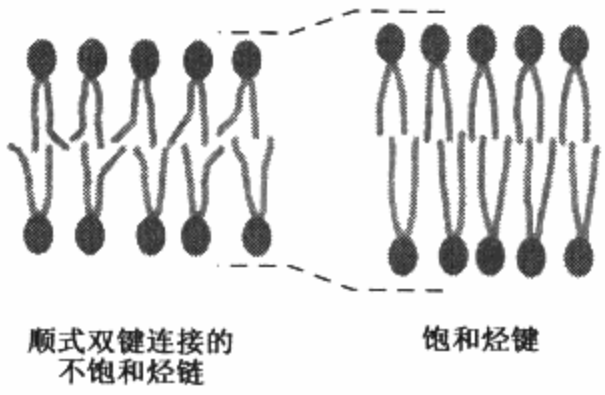


图 10-9　烃链中顺式双键的影响

双键使得更难将那些链聚集在一起，因此使得脂双层更难被固定。此外，因为不饱和脂分子的脂肪酸链伸展的分开，所以含有这些分子的双分子层要比那些专门由饱和脂分子组成的双分子层薄

不过许多细胞膜的脂双层不仅仅由磷脂分子专门组成，通常也包括胆固醇和糖脂。特别是真核生物的质膜包含有大量的**胆固醇**（图 10-10），最高到一分子磷脂对应一分子胆固醇。胆固醇分子增强了脂双层通透—屏障特性。它们在双分子层中使它们的羟基靠近磷脂分子的极性头部基团。在这个位置，它们刚性的类固醇环和那些靠近极性头部基团的羟基链相互作用（图 10-11）。通过减少磷脂分子羟基链前端的 CH_2 基团的活动性，胆固醇使得那个区域的脂双层形变可能降低，因此降低了小的水溶性分子在双分子层的通透性。尽管胆固醇具有降低脂双层流动性的趋势，但是在大多数真核细胞质膜中发现其中具有高浓度的胆固醇，这也能防止烃链聚在一起和结晶。通过这种方式约束了可能的相变。

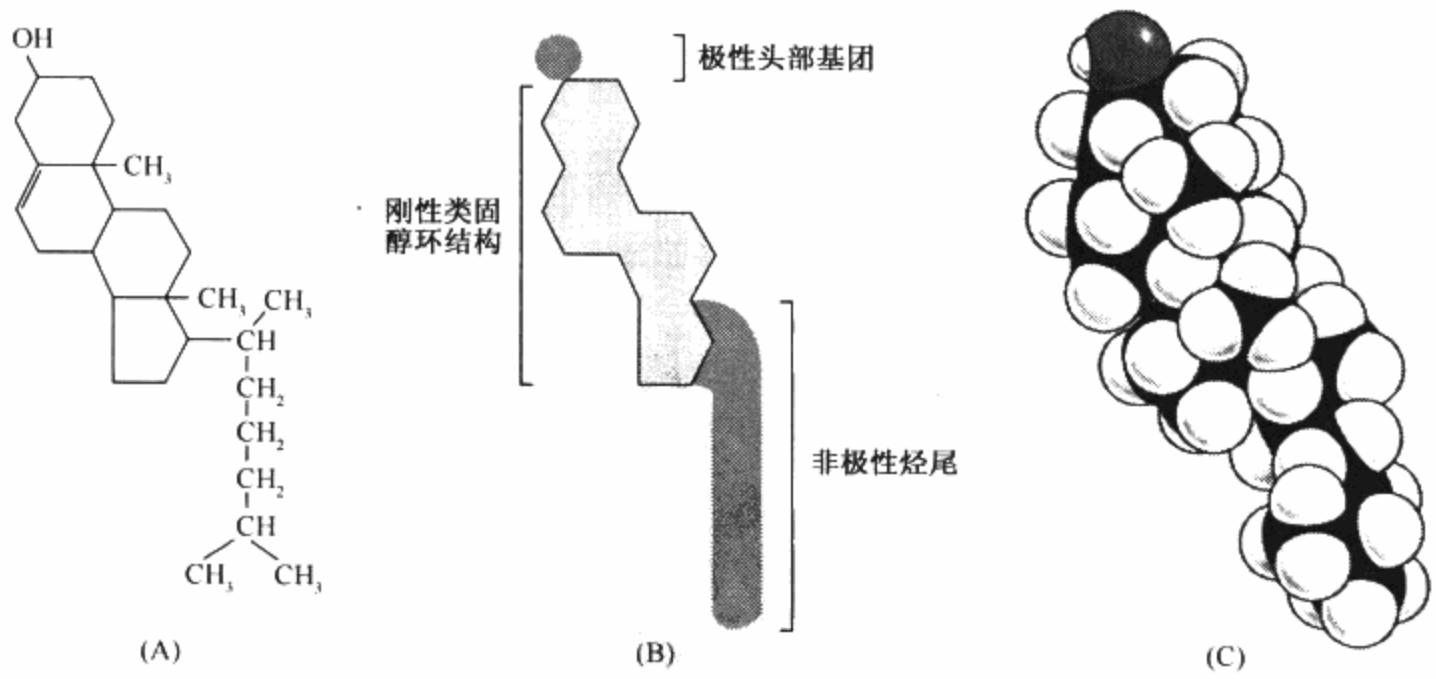


图 10-10　胆固醇的结构

胆固醇用以下几种形式表示：（A）结构式；（B）示意图；（C）空间结构模型

在表 10-1 中比较了几种生物膜中脂质成分。细菌质膜通常包括一种主要类型的磷脂分子成分，不含有胆固醇；通过其上的细胞壁可以增强膜的机械稳定性（图 11-17）。相反，大多

数真核细胞的质膜是更多样化的，不仅仅包含大量的胆固醇，也包含不同的磷脂分子的混合物。

在许多哺乳动物的质膜中以 4 种磷脂分子为主：卵磷脂、磷脂酰乙醇胺、磷脂酰丝氨酸和鞘磷脂。这些分子的结构显示在图 10-12 中。值得注意的是只有磷脂酰丝氨酸带有净负电荷，这个的重要性我们随后讨论；其他三种在生理的 pH 中是电中性的，带有一个正电荷和一个负电荷。在大多数膜中的脂质质量大多数由这 4 种脂分子共同构成（表 10-1）。其他的磷脂分子，像肌醇磷脂，虽然在数量上很少，但功能上非常重要。就像第 15 章描述的，肌醇磷脂在细胞信号通信中扮演了至关重要的角色。

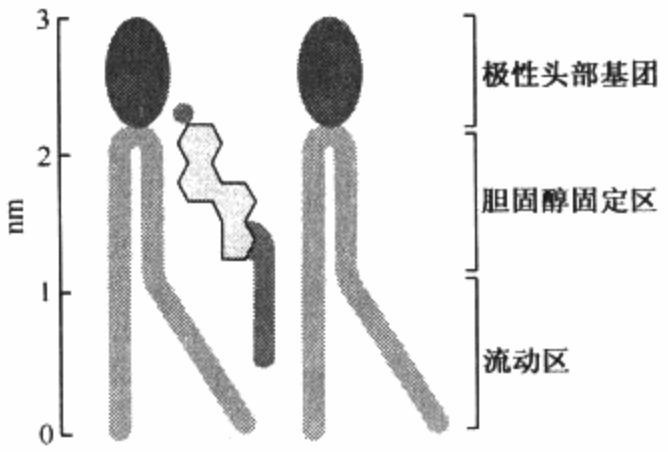


图 10-11　在脂双层中的胆固醇

示意图显示的是一个胆固醇分子在脂双层的一个单层中和两个磷脂分子相互作用

表 10-1　不同细胞膜大致的脂质成分

脂　质	占总脂质重量的百分比					
	活细胞的质膜	红细胞质膜	髓磷脂	线粒体（内膜和外膜）	内质网	*E. coli*
胆固醇	17	23	22	3	6	0
磷脂酰乙醇胺	7	18	15	25	17	70
磷脂酰丝氨酸	4	7	9	2	5	痕量
卵磷脂	24	17	10	39	40	0
鞘磷脂	19	18	8	0	5	0
糖脂类	7	3	28	痕量	痕量	0
其他	22	13	8	21	27	30

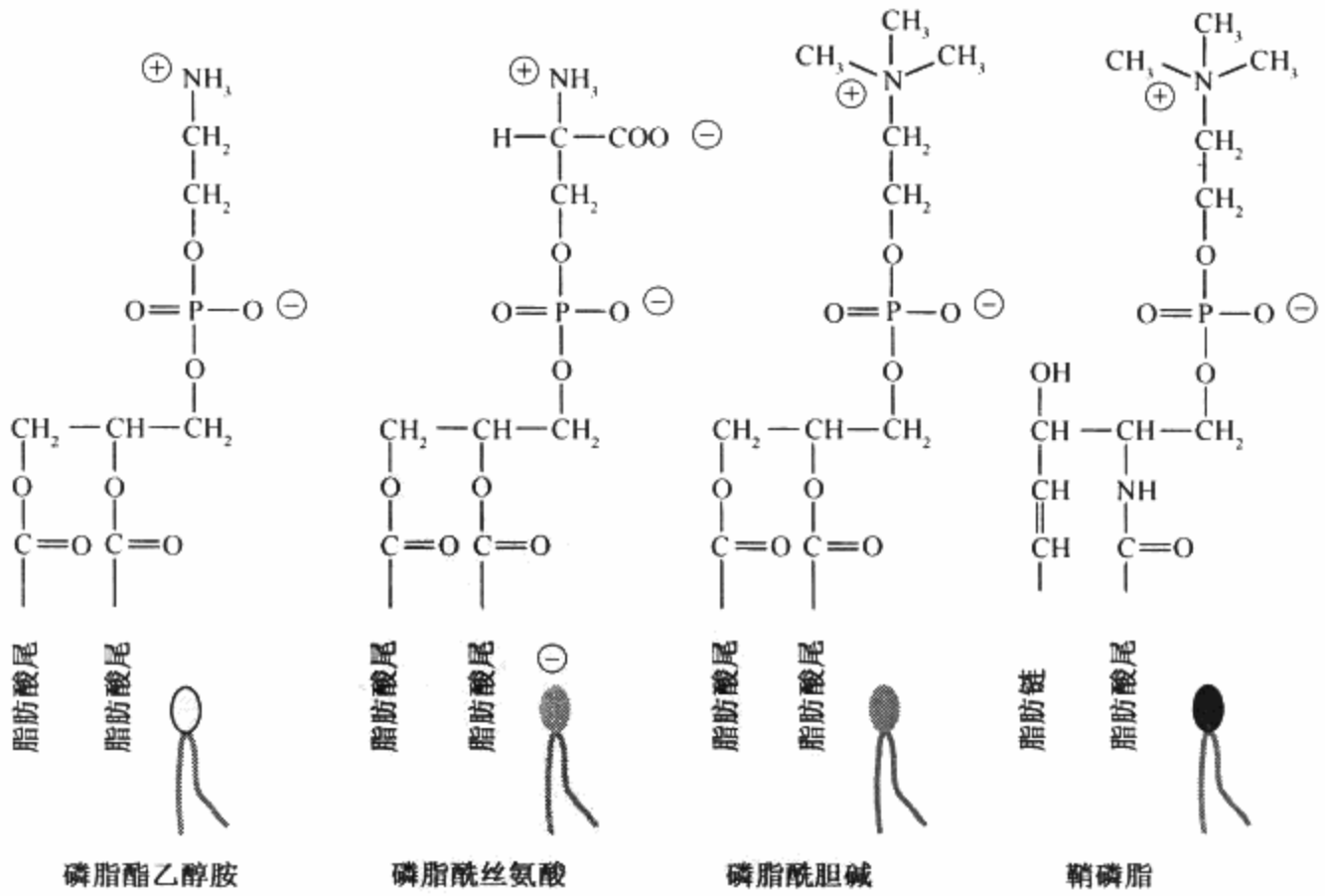


图 10-12　哺乳动物质膜中 4 种主要的磷脂

需要注意的是用不同的颜色表示的那些不同的头部基团。显示的所有脂类分子都源于甘油，除了鞘磷脂，它源于丝氨酸

人们可能会想知道为什么真核细胞的膜中包含如此多种类的磷脂，这些磷脂具有不同大小、形状和电荷的头部基团。一个人首先想要了解的是为什么人们认为膜质分子对于细胞膜中的蛋白质而言是二维溶剂，而水对于水溶液中的蛋白质而言是三维溶剂。一些膜蛋白只有在特定磷脂头部基团存在时才能行使功能，就像许多在水溶液中的酶需要特定的离子才有活性一样。此外，一些细胞质中的酶会和暴露在膜胞质这一面的脂分子头部基团结合，因此可以被补充和集中到特定的膜位点。

质膜包含富含有鞘磷脂、胆固醇和某些膜蛋白的脂质筏

细胞膜中大多数类型的脂分子随意地混在它们所处的脂单层中。相邻的脂肪酸尾部之间的范德华引力不能选择性地将同类分子的基团保持到一起。然而，对于某些脂类分子，如鞘脂类分子，具有长的饱和脂肪烃链，它们之间的范德华引力足够强，可以在一些微小的区域中将邻近的分子暂时地维持在一起。这种微小区域，或者称做**脂质筏**，鞘脂类分子被富集在其中。可以被认为是流动的脂双层中的瞬时的相分离。

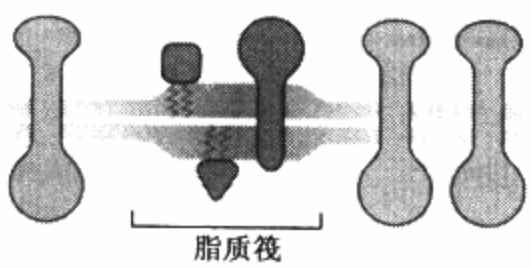


图 10-13　脂质筏

脂质筏是膜上小且特殊的区域，那里富集了一些脂分子（主要是鞘脂类和胆固醇）和蛋白质（绿色）。因为在脂质筏中某处的脂双层更厚，某些膜蛋白就聚集在那。关于脂质筏更详细的见图 13-63

通常认为动物细胞的质膜中包含有许多微小的脂质筏（直径 70nm 左右），其中具有丰富的鞘脂类分子和胆固醇。由于该处富集的烃链比大多数膜脂的脂肪酸链要长且直，因此脂质筏比双分子层的其他部分要厚（图 10-9），能够更好地容纳某些趋于集聚在该处的膜蛋白（图 10-13）。这样，通常认为脂质筏能帮助那些能将胞外信号转换成胞内信号的蛋白质，将那些蛋白质浓缩到小囊泡中或者帮助那些蛋白质集中到一起行使功能（讨论见第 15 章）。

在很大程度上，双分子层的一个单层中的脂分子的移动不受另一个单层中的脂分子的约束。然而在脂质筏中，鞘脂类分子在一个单层中的长烃链会和另一个单层中的长烃链相互作用。因此，脂质筏中两个单层通过它们的脂肪尾巴相连接。

脂双层的不对称性在功能上是重要的

在许多生物膜的脂质双分子层的两个单层中的脂质成分具有明显的不同。例如，在人红细胞膜中，几乎所有的头部基团中具有胆碱 [$(CH_3)_3N^+CH_2CH_2OH$] 的脂分子都在膜的外单层上，而几乎所有的具有氨基端的磷脂分子（磷脂酰乙醇胺和磷脂酰丝氨酸）都在膜的内

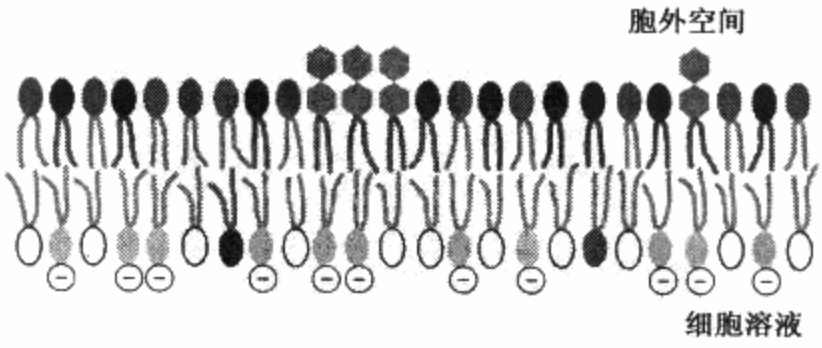


图 10-14　磷脂和糖脂在人红细胞中的不对称分布

用于磷脂分子头部基团的那些颜色在图 10-12 中有介绍。此外，糖脂用六边形极性头部基团（蓝色）画出。胆固醇（没显示出来）被认为在 2 个单层中平均分布

单层上（图 10-14）。因为带负电的磷脂酰丝氨酸定位在内单层，导致了双分子层的内外两层在电荷上具有显著的不同。我们将在第 12 章讨论膜结合的磷脂转运蛋白如何产生和维持脂分子的不对称性。

脂类分子的不对称在功能上具有重要意义。许多细胞质中的蛋白质结合到那些在脂双分子层的胞质层中发现的特定的脂分子头部基团上。例如，蛋白激酶 C（protein kinase C，PKC）能够被多种细胞外信号的应答所激活。它结合在质膜的胞质面，磷脂酰丝氨酸集中在那，酶的活性需要这种带负电的磷脂的存在。

在其他的研究案例中，脂分子的头部基团必须先被修饰，因此蛋白质结合位点在特定的时间和空间上被确立。以磷脂酰肌醇为例，这是一种含量较少的磷脂，这种磷脂集中在细胞膜的胞质面。一系列的脂质激酶可以将磷酸基团加到肌醇环的特定位置上。然后磷酸化的磷脂酰肌醇起着将特定蛋白质从细胞质中补充到膜上的作用。这种脂质激酶中一个重要的实例是磷脂酰肌醇激酶（phosphatidylinositol kinase，PI 3-kinase），它能被细胞外信号的应答激活，有助于将特定的细胞内信号蛋白补充到质膜的胞质面［图 10-15(A)］。相似的脂质激酶使得在细胞内膜上的肌醇磷脂磷酸化，然后有助于那些引导膜运输的蛋白质的补充。

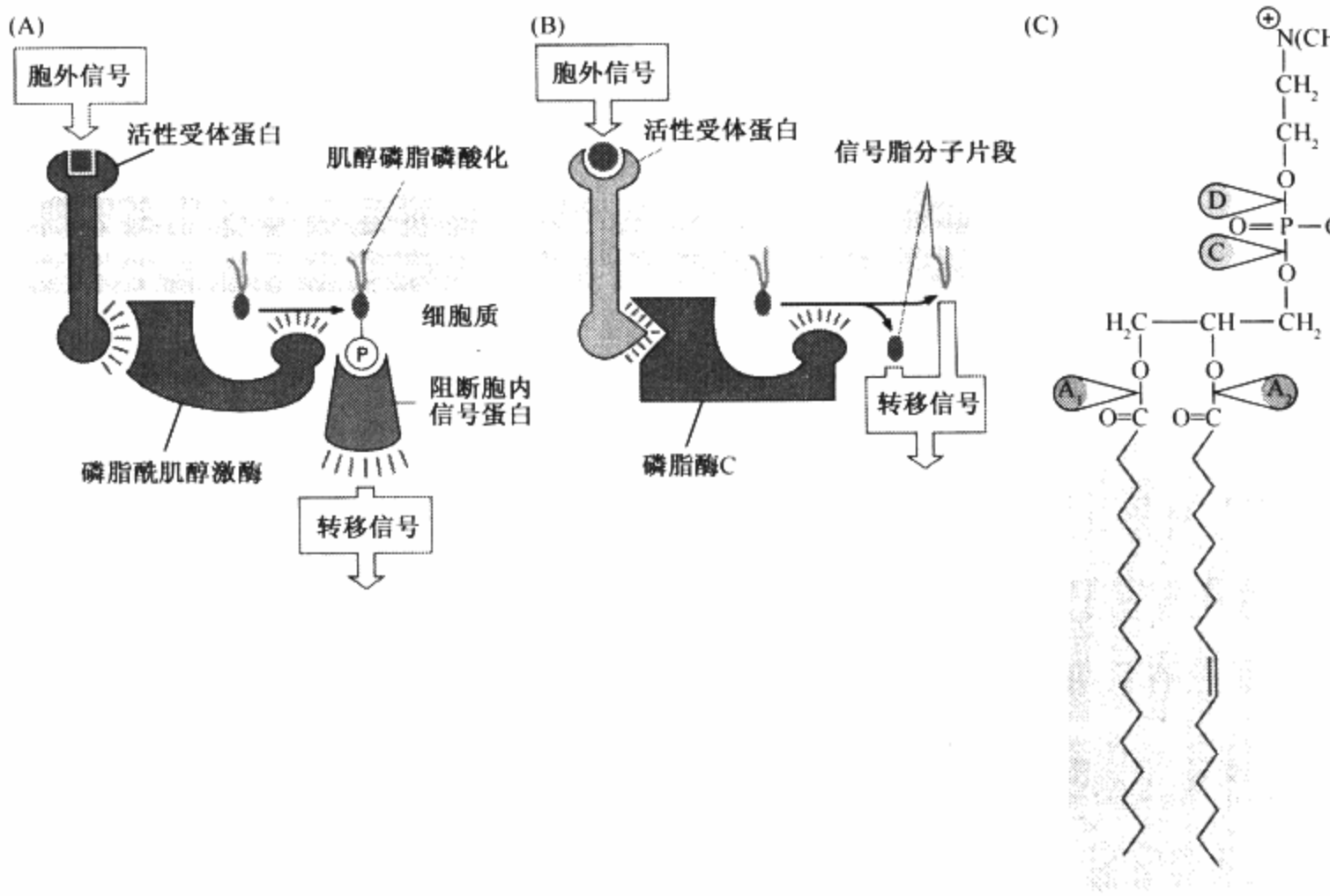


图 10-15　膜磷脂在细胞信号传递中的一些功能

（A）细胞外信号能激活 PI 3 激酶，这能使细胞膜中的肌醇磷脂磷酸化。随后多种细胞内信号分子和这些磷酸化的脂分子结合，因此被补充到膜上，在膜上信号分子能相互作用有助于将信号转移到细胞中。（B）其他细胞外信号活化磷酸酶切开磷脂。然后脂分子片段作为信号分子将信号转移到细胞中。（C）图解显示了不同类型的磷脂酶切开磷脂分子的位点。正如图中指出的，磷脂酶 A_1 和磷脂酶 A_2 切开酯键，而磷脂酶 C 和磷脂酶 D 切开磷脂键

质膜中的磷脂通常也采用另一种方式对胞外信号应答。质膜包含多种磷脂酶，这些酶能被胞外信号激活后可以劈开特定的磷脂分子，产生的这些分子的碎片可以作为具有短时活性

的胞内信使［图 10-15(B)］。例如，磷脂酶 C 可以将质膜胞质层中的一个肌醇磷脂劈成两段，其中一个保持在膜中，帮助激活蛋白激酶 C，而另外一半被释放到胞质中，刺激内质网的 Ca^{2+} 释放（图 15-36）。

动物利用质膜上的磷脂不对称性来区别活细胞和死细胞。当动物细胞经历程序性死亡或凋亡时（讨论见第 17 章），通常被限制在质膜脂类双分子层的胞质单层中的磷脂酰丝氨酸，能够被快速的转运到胞外单层。暴露在细胞表面的磷脂酰丝氨酸会作为一个信号来引起周围的细胞，如巨噬细胞等，来吞噬这死亡的细胞并消化它。这种磷脂酰丝氨酸转运现象的产生基于以下两种机制：

1. 通常能将磷脂酰丝氨酸从非胞质单层转运到胞质单层的磷脂酰丝氨酸转运蛋白失活了。
2. 一种能够将磷脂酰丝氨酸在膜的两个单层之间双向非特异转运的酶被激活。

糖脂在所有质膜表面都有发现

在膜上分布最不对称的脂分子是含糖的脂分子，叫做糖脂。这些分子在脂双层的非细胞质单层上被发现，在那儿这些分子被认为优先被分隔成脂质筏。这些糖脂分子通过糖基彼此之间的氢键或它们饱和烃链之间的范德华力趋于自联合。糖脂在双分子层中的不对称分布源于在高尔基体内腔中糖基添加到脂分子上的过程，从拓扑学上看高尔基体的内腔等价于细胞的外表面（讨论见第 12 章）。在质膜中，糖基暴露在细胞表面（图 10-14），在细胞和周围环境的相互作用中，它们扮演了重要角色。

糖脂大概存在于所有动物细胞的质膜中，在那些质膜中糖脂通常占质膜外层的脂分子的

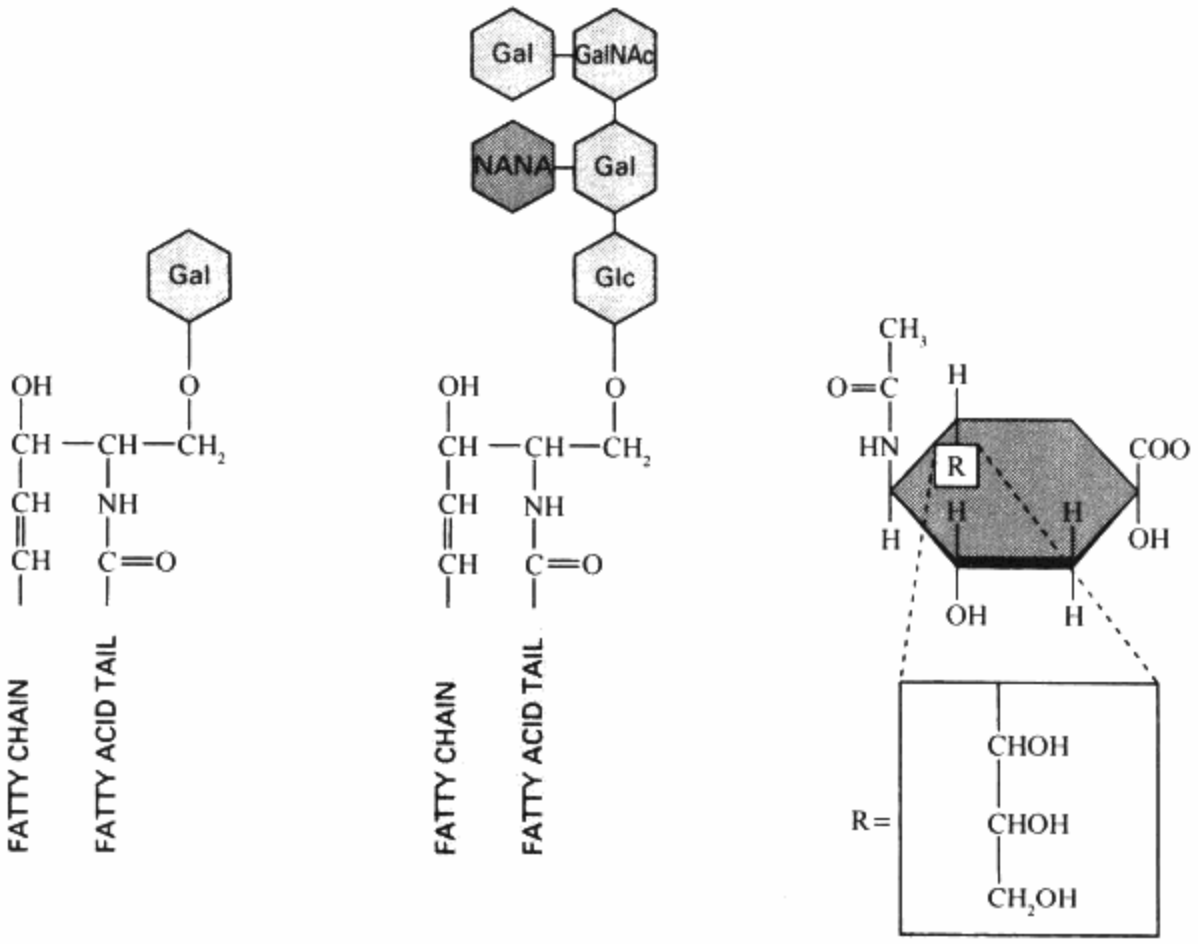


图 10-16　糖脂分子

（A）半乳糖脑苷脂被称作中性糖脂，因为形成头部基团的糖基是不带电的。（B）神经节苷脂总是包含一个或多个唾液酸残基（也叫做 *N*-acetylneuraminic acid，NANA），结构显示在（C）中。无论细菌中还是植物中，几乎所有的糖脂都来源于甘油；在动物细胞中大多数磷脂都从丝氨酸中产生，对于鞘磷脂也是这样（图 10-12）

5%。在一些细胞内膜中也发现有糖脂。作为复杂的糖脂，**神经节苷脂**包含有一个或多个带有唾液酸残基的寡糖，这给神经节苷脂一个净负电荷（图 10-16）。超过 40 个神经节苷脂已经得到鉴定。这些在神经细胞的脂膜中含量丰富，神经节苷脂占那些地方脂类总质量的 5%～10%；在其他类型的细胞中也发现有少量的神经节苷脂存在。

关于糖脂具有哪些功能的线索可能来自于它们的定位。例如，在上皮细胞的质膜中，糖脂被限制在暴露在外的顶部表面上，在那里它们可能有助于保护细胞膜抵抗苛刻的条件（如低 pH 和有降解能力的酶类）。带电的糖脂，如神经节苷脂，可能对于它们的电效应很重要：它们的存在改变了跨膜的电场和膜表面的离子浓度，尤其是 Ca^{2+} 的浓度。糖脂也被认为在细胞识别过程中具有作用，在细胞间的黏附过程中糖脂和糖蛋白上的糖基都能结合到膜上的糖基结合蛋白（凝集素）上（讨论见第 19 章）。令人惊讶的是，那些缺失神经节苷脂的突变小鼠并没有显示明显的畸形，尽管在实验中雄性小鼠不能正常的转运睾酮并因此不育。

无论它们本身功能如何，有些糖脂为某些细菌毒素提供了入口。例如，神经节苷脂 G_{M1}（图 10-16）充当了那些能引起霍乱痢疾的细菌毒素的细胞膜表面受体。霍乱毒素只能结合并进入那些表面具有 G_{M1} 的细胞，包括肠上皮细胞。它进入细胞后会导致细胞内 cAMP 的浓度的急剧增长（讨论见第 15 章），这接着引起大量的 Na^+ 和水流到肠道内。

小结

生物膜由连续的双重脂类分子双层组成，膜蛋白嵌于其中。脂双层有流动性，独立的脂类分子快速分散在它们所处的单层之内。膜脂类分子是两性分子。其中最多的是磷脂。置于水中时它们能自组装成双分子层，如果被撕裂后还能重新密封形成密封腔。

膜脂类分子有磷脂、胆固醇和糖脂三大类。内外单层的脂类成分不同，表现为细胞膜两面具有不同的功能。脂类分子不同的混合物被发现在不同类型细胞以及单个真核细胞不同的膜上。某些膜结合酶需要特定的脂类头部基团（lipid head group）来行使功能。某些脂类分子的头部基团形成特定胞质蛋白的停靠位点。有些胞外信号通过膜受体蛋白激活磷脂酶，磷脂酶切断脂膜中特定的磷脂分子，因此产生的片段作为胞内信号分子起作用。

膜蛋白

虽然生物膜基本结构由脂双层提供，膜的大部分具体功能都由膜蛋白执行。因此就是这些蛋白质将细胞典型的功能特性赋予细胞中的各类膜。相应地，膜内蛋白质的数量和类型都是高度易变的。髓磷脂的膜作为神经细胞轴突的主要电子绝缘材料，蛋白质占该膜质量的 25%不到。相反，涉及到 ATP 生产的膜（如线粒体和叶绿体的内膜），大约 75%是蛋白质。一典型的细胞质膜介于两者之间，蛋白质含量大约占膜质量的 50%。

由于脂质分子比蛋白质分子小，膜上脂质分子总是比蛋白质分子要多。在蛋白质含量占膜质量的 50%的膜上，大约每 50 个脂质分子对应一个蛋白质分子。和膜脂质相似，膜蛋白上常有低聚糖链附着，糖链朝向细胞外部。因此细胞存在富含糖类的外表面，形成了细胞衣，正如我们接下来要讨论的。

膜蛋白可以通过多种方法和脂双层联在一起

不同的膜蛋白通过不同方法和膜相结合，如图 10-17 所示 。许多蛋白质扩散在整个

脂双层，每一边都有一部分（图 10-17 中的例子①、②、③）。像它们的脂质邻居，这些**跨膜蛋白**是两性分子，具有疏水区域和亲水区域。它们的疏水区域通过膜，和脂双层内部脂质分子的疏水尾巴结合，和水隔离开。它们的亲水区域在膜的两边暴露于水中。通过和插入脂双层的胞质单层的脂肪酸链共价连接，某些跨膜蛋白的疏水性能得到提高（图 10-17 例①）。

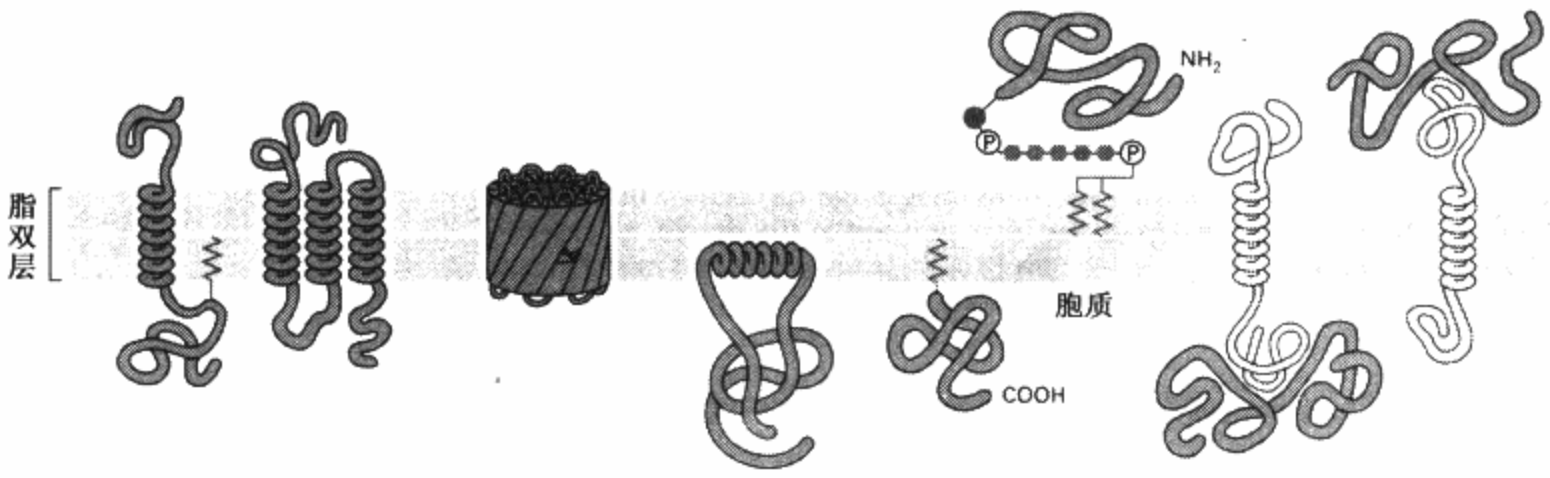


图 10-17　膜蛋白和脂双层连接的多种方式

通常认为大多数跨膜蛋白以以下形式穿过双分子层：①单个的 α 螺旋。②多个 α 螺旋。③成一个卷的 β 片段（β 桶）。这些单次跨膜或多次跨膜的蛋白质和插入在胞质单层的脂肪酸链有一共价结合。其他膜蛋白只暴露在膜的一侧。④这其中一些蛋白质通过一个两性的部分处于和脂双层的胞质单层中的 α 螺旋锚在胞质面。⑤其他蛋白质单独地通过在胞质单层中的脂分子链（脂肪酸链或异戊烯基团）共价连接到双分子层上。⑥或者通过寡糖衔接物连接非胞质层中的磷脂酸肌醇。⑦和⑧最终，许多蛋白质只是通过和其他膜蛋白的非共价相互作用连接到膜上。⑨中结构形成的方式在图 10-18 中阐明。膜蛋白如何和脂双层连接的细节在第 12 章讨论

其他完全定位在细胞质中的膜蛋白，通过暴露在蛋白质表面的两性的 α 螺旋和脂双层的胞质单层连接（图 10-17 例④）或者通过脂肪酸链或异戊二烯基团之类的脂质链共价连接（图 10-17 例⑤和图 10-18）。不过其他那些完全暴露在细胞外表面的膜蛋白，只能通过和细胞质膜外层的磷脂酰丝氨酸共价连接（通过一特定低聚糖）和脂双层结合（图 10-17 例⑥）。

图 10-17 例⑤中和脂质连接的蛋白质是在细胞质中合成的可溶性蛋白质，随后通过和脂质基团的共价吸附直接连到膜上（图 10-18）。但是图 10-17 例⑥中的蛋白质是在内质网中合成的单通跨膜蛋白。当仍然在内质网中时，蛋白质的跨膜片段被劈开，加入**糖磷脂酰肌醇（GPI）**锚，通过这个锚引导蛋白质绑定到细胞膜非细胞质的那一面（讨论见第 12 章）。使用一种叫做糖磷脂酰肌醇特异性磷脂酶 C，能够容易地辨别那些靠 GPI 锚绑定到细胞质膜上的蛋白质。这种酶将蛋白质从锚上切下，将蛋白质从膜上释放。

一些膜蛋白从不延伸到脂双层的疏水内部；它们通过和其他膜蛋白的共价作用绑定到膜的每一面（图 10-17 例⑦和例⑧）。很多这种类型的蛋白质可以通过相对温和的提取过程从膜上释放出来，例如，暴露在离子浓度极高或极低的溶液或极端 pH 的溶液中，这虽然会干扰蛋白质与蛋白质间的相互作用，但是能保持脂双层的完整；这些蛋白质被称作**外周膜蛋白**。大多数跨膜蛋白脂质基团保持在脂双层中，还有一些通过和其他蛋白质牢固地结合保持在膜上，不能通过这些方法释放。这些蛋白质被称作**内在膜蛋白**。

膜蛋白如何和脂双层连接反映了蛋白质功能。只有跨膜蛋白能够在脂双层的两边执行功能或运输分子通过膜。细胞表面受体是跨膜蛋白，在细胞外部结合信号分子后在质膜反面产生不同的细胞内信号。与之相对的是，只在脂双层的单面发挥功能的蛋白质通常只和脂质单

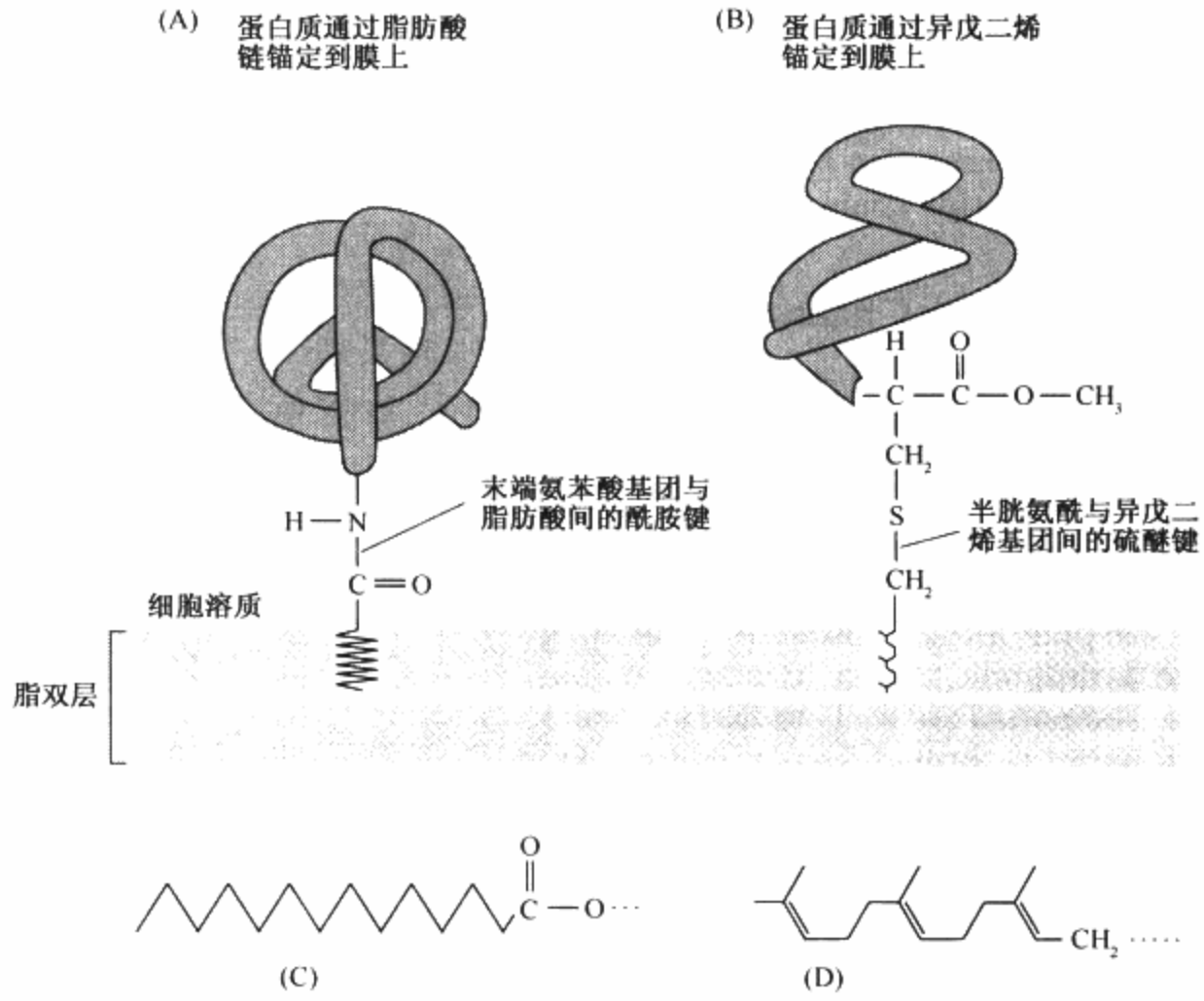


图 10-18　由脂肪酸链或异戊二烯结合的膜蛋白质

每类脂分子的共价结合有助于水溶性蛋白质在细胞质中合成后定位到膜上。（A）一个脂肪酸链（肉豆蔻酸）通过氨基连接到甘氨酸的 N 末端。（B）一个异戊二烯基通过一个二硫键连接到位于蛋白质 C 端的 4 个残基开头的半胱氨酸残基上。在异戊烯化后，最末端的 3 个氨基酸被切下来，再插入膜之前蛋白质新的 C 端被甲基化。棕榈酸，一种 18 个碳原子的饱和脂肪酸，也能通过二硫键形式和内部的半胱氨酸侧链相连到某些蛋白质上。这种允许蛋白质补充到膜上的修饰通常是不可逆的。这两种脂分子锚的结构显示如下：（C）一个十四碳饱和脂肪酸的锚。（D）一个十五碳不饱和烃链的锚

层或某一面的蛋白质域专一结合。例如，涉及细胞内信号的某些蛋白质，通过一个或多个脂质基团和质膜细胞质那一面共价结合。

大多数跨膜蛋白借助 α 螺旋构象的多肽链穿过脂双层

一个跨膜蛋白在膜中总是有一个独特的方向。这既反映蛋白质在内质网中的合成及插入到脂双层中不对称的方式，也反映了蛋白质胞质和非胞质区域功能的不同。这些区域被多肽链的跨膜片段所分隔，这些跨膜片段联系着脂双层的疏水环境，主要由非极性侧链的氨基酸残基组成。因为肽键本身是极性的而且双分子层中不存在水，所以双分子层中所有的肽键彼此之间形成氢键。如果多肽链穿过双分子层是形成规则的 α 螺旋，这时肽键之间的氢键是最优化的，这也是大多数多肽链跨膜片段穿过双分子层的形式（图 10-19）。

在**单次跨膜蛋白质**中，多肽只穿过膜一次（图 10-17 例①），而在**多次跨膜蛋白质**中，多肽链穿过膜多次（图 10-17 例②）。对于要满足形成氢键的需要，脂双层中肽键可供用的方式是多肽链多次跨膜，跨膜片段按照 β 折叠排列成封闭的桶形结构（也叫做 β 桶，图 10-17 例③）。这种多次跨膜的结构在孔道蛋白（porin protein）中被发现，我们随后会讨论。这种无水环境中最大限度的形成氢键的动力也意味着进入脂双层的多肽链很可能在改变方向之前就完全穿过了脂双层，因为链的弯曲会导致正常氢键的相互作用的损失。

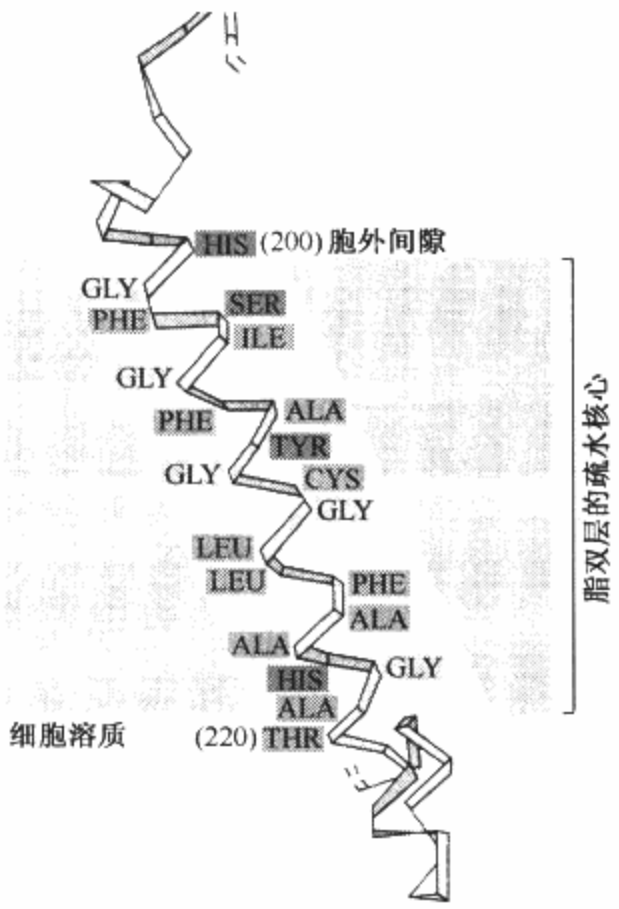


图 10-19　以 α 螺旋穿过脂双层的跨膜多肽的片段

只显示了多肽链的 α 螺旋的碳骨架，其中疏水氨基酸用绿色和黄色显示。图中显示的多肽片段是图 10-38 所描绘的细菌光合作用中心的一部分，这个结构通过 X 射线衍射测定

由于跨膜蛋白很难结晶，相应地，这类蛋白质很少被 X 射线晶体学完全研究清楚。几乎所有的折叠的三维结构都是不确定的。然而，DNA 克隆和测序技术已经显示了大量的跨膜蛋白的氨基酸序列，这通常可根据蛋白质序列的分析来预言蛋白质中跨过脂双层的多肽链部分。这些包含 20～30 个高疏水性的氨基酸的片段足够长，可以形成跨过脂双层的 α 螺旋，这些片段通常依靠亲水性图来鉴定（图 10-20）。从这些图中可以预知一个生物体产生的蛋白质中跨膜蛋白的比例。例如，对于出芽酵母，它的全基因组的核酸序列是已知的，大约 20%的蛋白质被鉴定为跨膜蛋白，这也强调了膜蛋白功能上的重要性。亲水性图不能鉴定 β 折叠的跨膜片段，因为作为 β 链穿过脂双层只需要 10 个或更少的氨基酸，而且这些氨基酸侧链只需要每隔一个是疏水的就可以。

某些 β 折叠形成大的跨膜通道

那些具有 β 折叠作为跨膜片段的多次跨膜蛋白比那些具有 α 螺旋跨膜片段的蛋白质来说，具有更大的刚性也更容易趋于结晶。因此许多这种蛋白质的结构已经通过 X 射线晶体学测定出来。β 链的数量变化范围很宽，可以少到 8 条链也可以多到 22 条链（图 10-21）。

β 折叠蛋白在线粒体、叶绿体和许多细菌的外部膜中含量丰富。有些是成孔蛋白，产生充满水的通道，允许特定的亲水溶质通过细菌的外部膜穿过脂双层。孔道蛋白是被充分研究的例子（图 10-21 例③），孔道蛋白折叠由 16 条链的反向平行的 β 区组成，那足够大到形成圆柱形结构。极性侧链排列成内部的水相通道，而非极性侧链突出于折叠的外部和脂双层的疏水核心相互作用。多肽链的环状结构通常突出进入通道的内腔，收缩通道使得只有特定的溶质才能通过。因此有些孔道蛋白具有很好的选择性：如麦芽糖孔蛋白（maltoporin）就优先允许麦芽糖或低聚麦芽糖穿过大肠杆菌的外部膜。

FepA 蛋白是这类转运蛋白中更复杂的一个例子（图 10-21 例④）。它运输铁离子通过细菌外部膜。它的大圆桶结构由 22 个 β 链组成，在圆筒的内部被一个大的球状区域充满。铁离子结合在那个区域，通常认为运输铁离子通过膜时要经历大的构象变化。

并不是所有的 β 桶状蛋白都是转运蛋白。有些蛋白质由于氨基酸侧链突出到桶状结构的中心，形成更小的桶状结构。这些蛋白质作为受体或酶行使功能（图 10-21 例①和例②）；其中桶状结构主要被用来作为刚性锚将蛋白质维持在膜上，以及确定胞质中环状区域的方向来形成特殊的胞内分子的结合位点。

尽管 β 桶状结构能适合多种用途，但这种结构很大程度上局限在细菌、线粒体和叶绿体的外部膜上。真核细胞和细菌脂膜中大多数的多次跨膜蛋白形成的是跨膜 α 螺旋。这些蛋白质中的这些螺旋能够彼此之间相对滑动，通过允许蛋白质的构象变化来开放或关闭离子通道、运输溶质或将胞外信号转换成胞内信号。相反，在 β 桶状结构蛋白中，每个 β 链和它相邻的

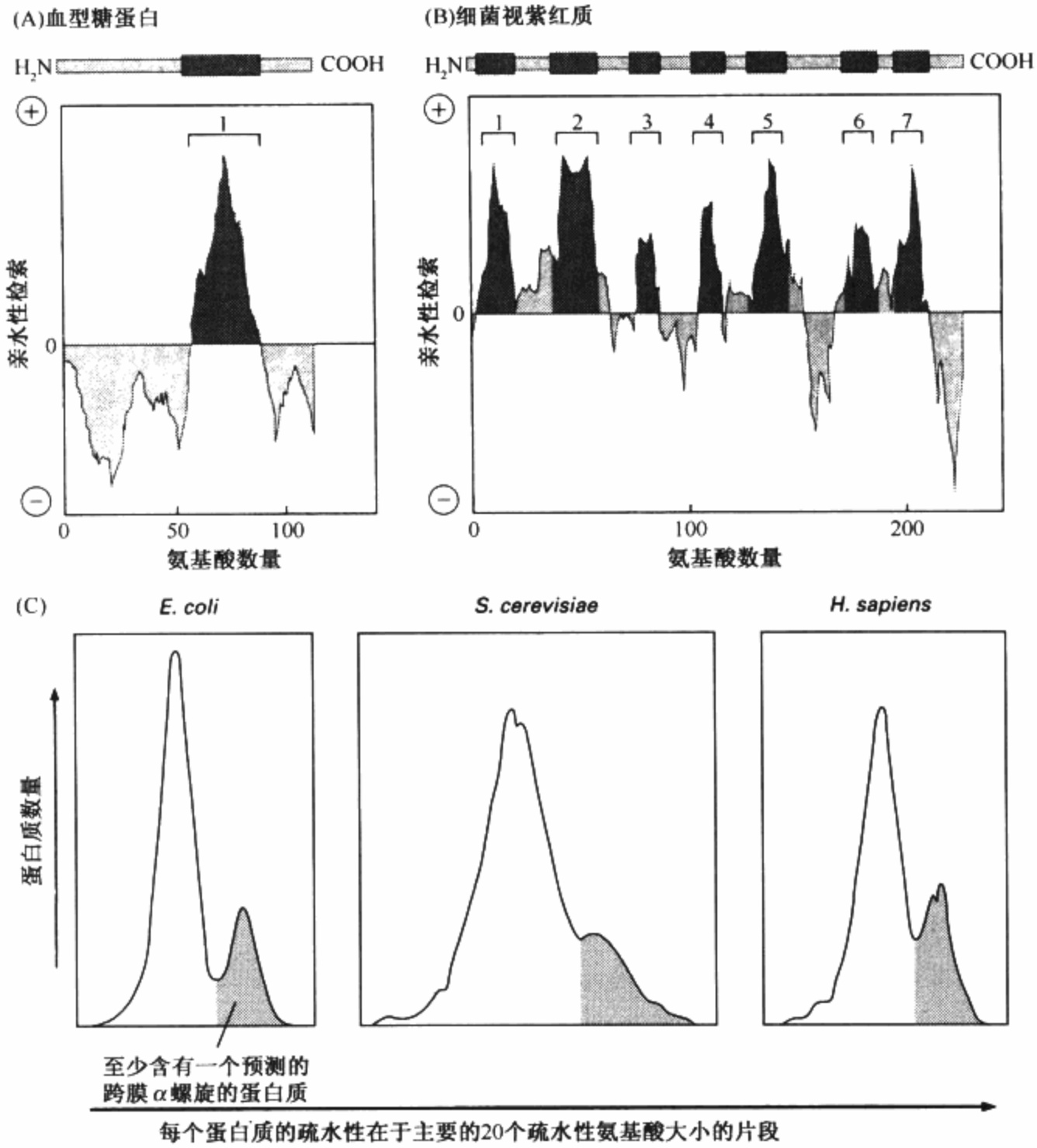


图 10-20　使用亲水性图来确定多肽链中可能的 α 螺旋跨膜片段的位置

对于将多肽链的连续片段从非极性溶剂转移到水中所需的自由能，可以根据每个片段所含氨基酸成分来计算，计算所用数据可以从一些典型化合物获得。这种计算通常由对固定大小的片段的计算组成（通常 10～20 个氨基酸），从链的每个连续的氨基酸开始。根据片段在链中位置的功能，片段的“亲水性检索”被绘图在 y 轴上。正值意味着转移到水中时需要的自由能（也就是说片段是疏水的），所指定的值就是所需能量数量的指标。亲水性检索的峰显示了氨基酸序列中疏水片段的位置 。(A) 和 (B) 显示的是后面讨论的两种膜蛋白。(A) 血型糖蛋白有 1 个单个的跨膜 α 螺旋以及在亲水性图有 1 个相应的峰。(B) 细菌视紫红质有 7 个跨膜 α 螺旋，在亲水性图有 7 个相应的峰。(C) 在 *E. coli*、*S. cerevisiae* 和人的基因组中对膜蛋白所占比例的预测。绿色区域显示的是至少含有一个预测的跨膜螺旋的蛋白质部分。*E. coli* 和 *S. cerevisiae* 的曲线代表了全基因组；人蛋白质的曲线代表的是不完全的基因组；在每种情况中，曲线下的面积是被分析的基因所占的比例［(A) 摘自 D. Eisenberg, *Annu. Rev. Biochem.* 53：595～624，1984；(C) 摘自 D. Boyd et al.，*Protein Sci.* 7：201～205，1998］

通过氢键刚性结合，这使得它不太可能发生自身桶状的构象变化。

许多膜蛋白被糖基化

动物细胞中大多数跨膜蛋白被糖基化。就像糖脂中一样，这些糖基是在内质网和高尔基体的内腔中加上的（讨论见第 12 和第 13 章）。因为这样，寡糖链总是出现在膜的非胞质面。膜两面蛋白质之间其他的区别由细胞质的还原环境产生。这种环境降低了膜胞质侧的链内和

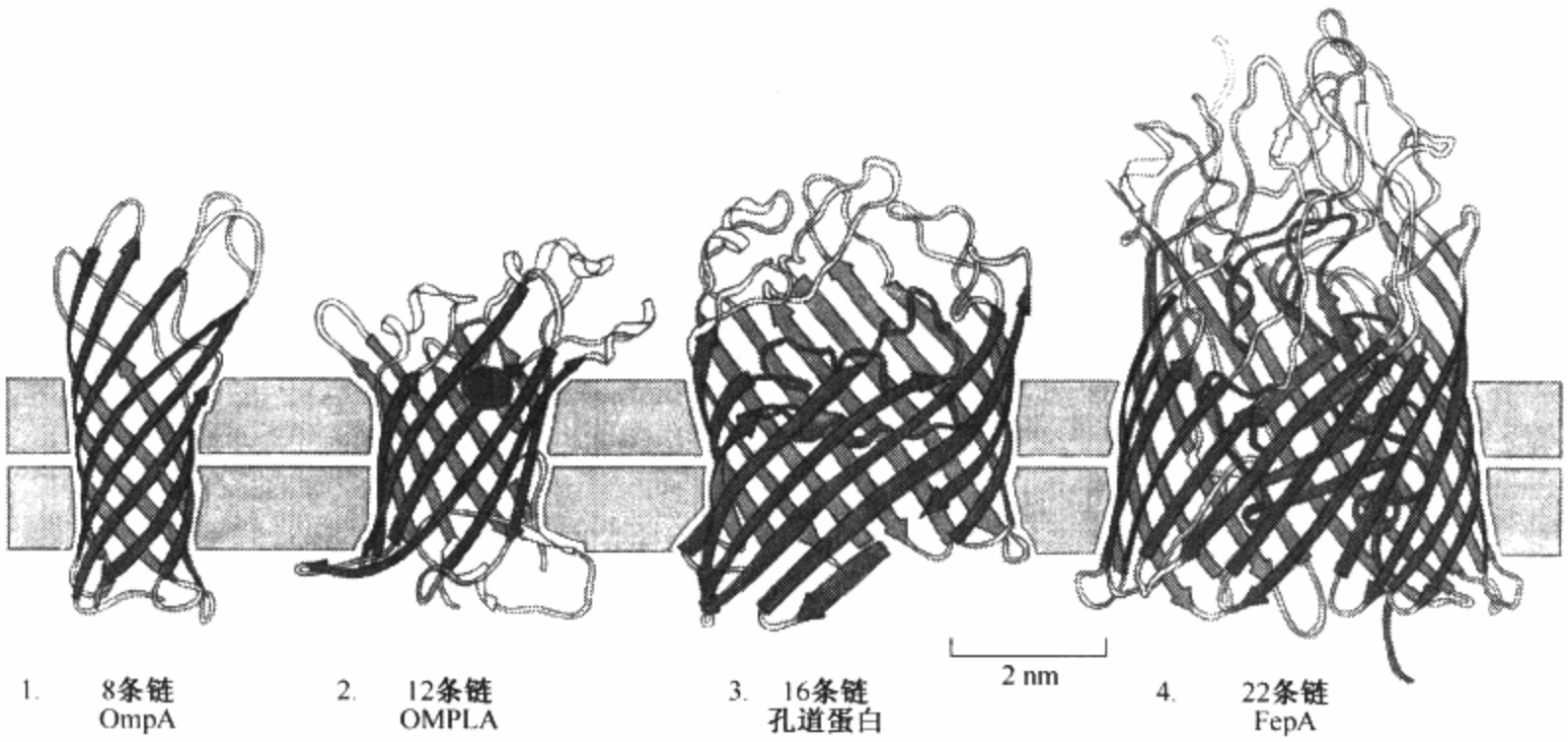


图 10-21　不同数量的β链的形成的β桶形结构

1. 大肠杆菌 OmpA 蛋白（8 条β链），作为一个噬菌体的受体起作用。2. 大肠杆菌 OMPLA 蛋白（12 条β链），是一个水解脂肪分子的脂肪酶。催化酶反应的氨基酸（用红色显示的），突出在桶状结构外侧的表面。3. 细菌 *Rhodobacter capsulatus* 中的一种孔道蛋白，形成了穿过外膜的充水的孔（16 条β链）。通道的直径受那些突出在通道中的茎环（用蓝色显示的）限制。4. 大肠杆菌 FepA 蛋白（22 条β链），转运铁离子。桶状结构的内部被一个球蛋白区域（用蓝色显示的）所充满，在那包含了铁结合位点。现在认为通过变化这个区域的构造来转运所结合的铁离子，但是这种变化的分子细节还不知道

链间二硫键形成的可能性。这些链内和链间的键会在非胞质侧形成，这些键在稳定多肽链的折叠结构或帮助多肽链和其他多肽链结合方面起着重要作用（图 10-22）。

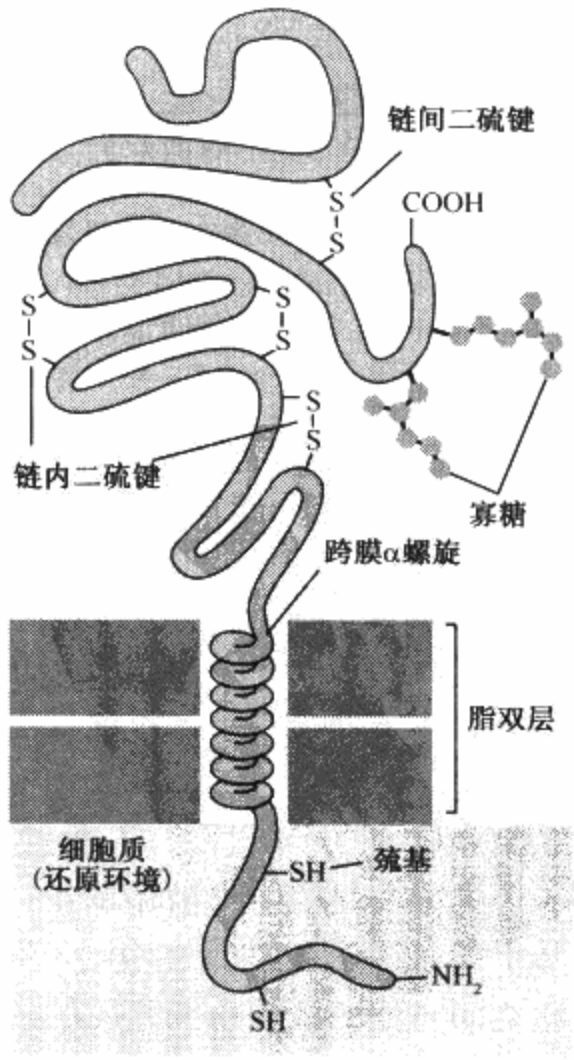


图 10-22　单次跨膜蛋白

需要注意的是多肽链以右手α螺旋穿过脂双层，寡糖链和二硫键都在膜的非胞质面。由于细胞质主要是还原环境，细胞中的那些基团保持还原（—SH）形式，因此蛋白质胞质区的巯基通常不能形成二硫键

用去污剂能溶解并纯化膜蛋白

通常，跨膜蛋白只能通过那些能破坏疏水缔合和脂双层的试剂来增加可溶性。对于膜生物化学家来说最有用的试剂是去污剂，这是一些小的两性分子，在水中趋于形成胶束（图 10-23）。和膜混合后，去污剂的疏水末端结合到膜蛋白的疏水区上，从那替代了脂分子。因为去污剂分子的另外一端是极性的，这种和膜蛋白的结合导致了形成的去污剂-蛋白质复合物被带到溶液中（尽管还有些脂分子可能还附着在蛋白质上）（图 10-24）。去污剂的极性（亲水）末端可以是离子化的，如在十二烷基磺酸钠（SDS）中；也可以是非离子化的，如 Triton 去污剂。这两种通用的去污剂的结构见图 10-25。

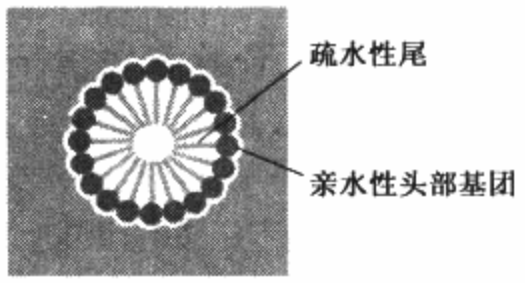


图 10-23　在水中一个去污剂胶囊，显示的是横截面

因为去污剂有极性和非极性末端，去污剂分子是两性分子。因为它们是楔形分子，所以相比形成双分子层而言，它们更容易形成胶囊（图 10-4）

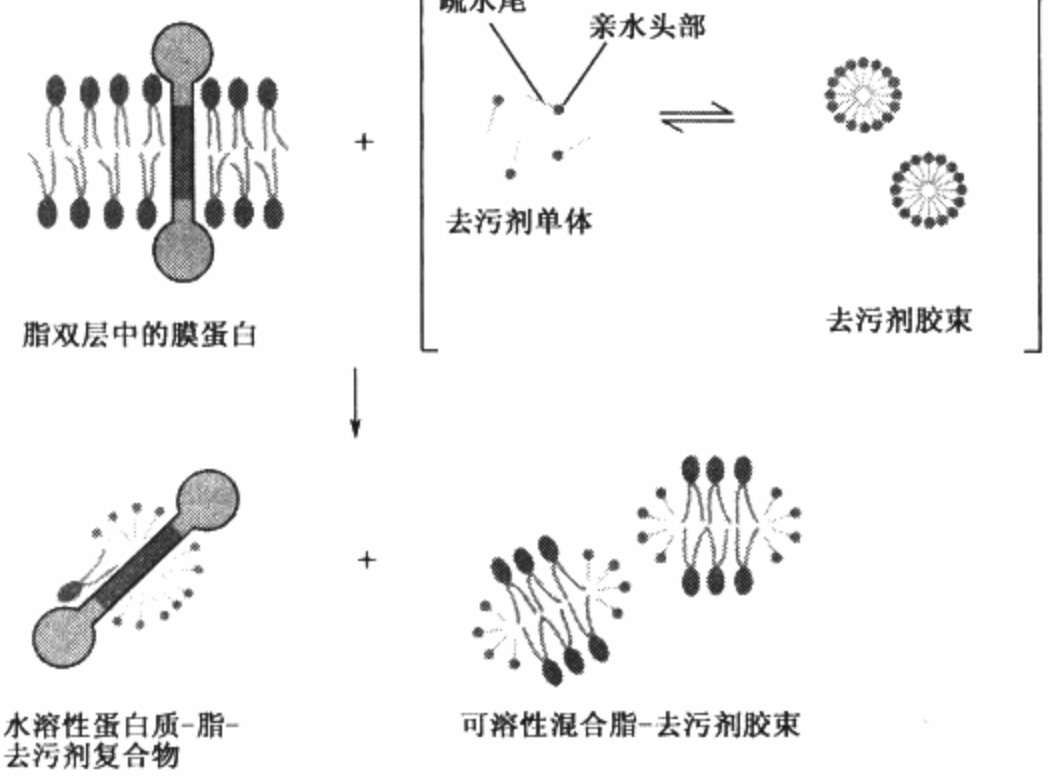


图 10-24　用中性去污剂溶解膜蛋白

去污剂破坏脂双层，通过形成蛋白质-脂-去污剂复合物将蛋白质带到溶液中。膜中的磷脂分子也能通过去污剂增溶

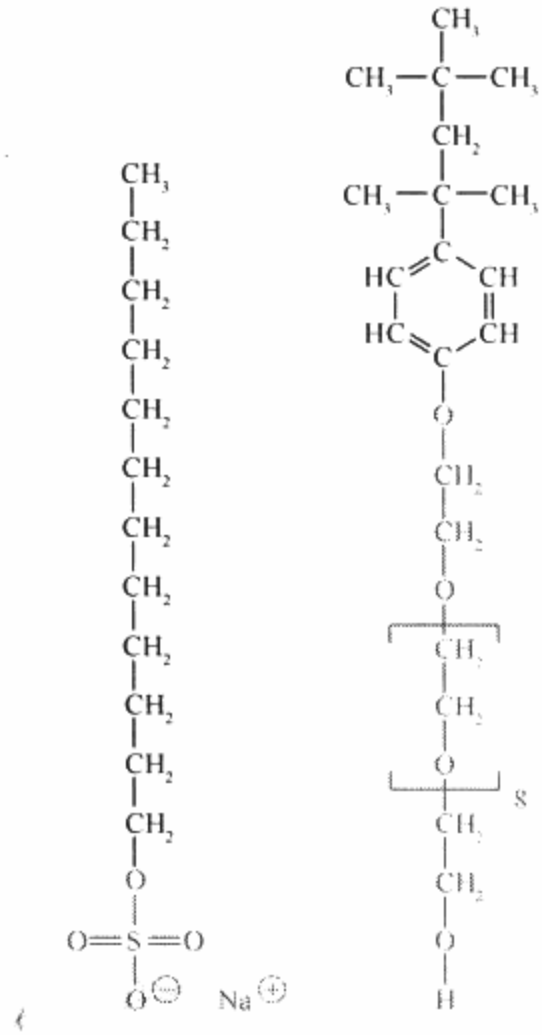


图 10-25　两种常用的去污剂的结构

十二烷基磺酸钠（SDS）是阴离子去污剂，Triton X-100 是非离子型去污剂。每种去污剂的疏水部分都用绿色显示，亲水部分用蓝色显示。Triton X-100 中括在一起的部分是一个 8 次重复

使用 SDS 之类的强力离子型去污剂，即使疏水性再强的膜蛋白也能被溶解。这使得可以用 SDS-PAGE 来分析这些蛋白质（讨论见第 8 章），这个过程革新了膜蛋白的研究。如此强

力的去污剂通过结合那些蛋白质内部的疏水中心来将蛋白质打开（使蛋白质变性），因此得到的蛋白质是无活性的也不能用于功能的研究中。不过蛋白质在 SDS 变性后很容易被纯化。在有些研究中，除去去污剂后纯化好的蛋白质还能复性，恢复功能上的活性。

通过使用一些温和的去污剂，许多疏水性膜蛋白能被增溶，然后可在保有活性的情况下得到纯化。像 Triton X-100 之类的去污剂可以覆盖在蛋白质跨膜片段上。通过这种方法，具有功能活性的膜蛋白系统能够从纯化后的成分中还原，这提供了分析这些蛋白质系统功能的有力手段（图 10-26）。

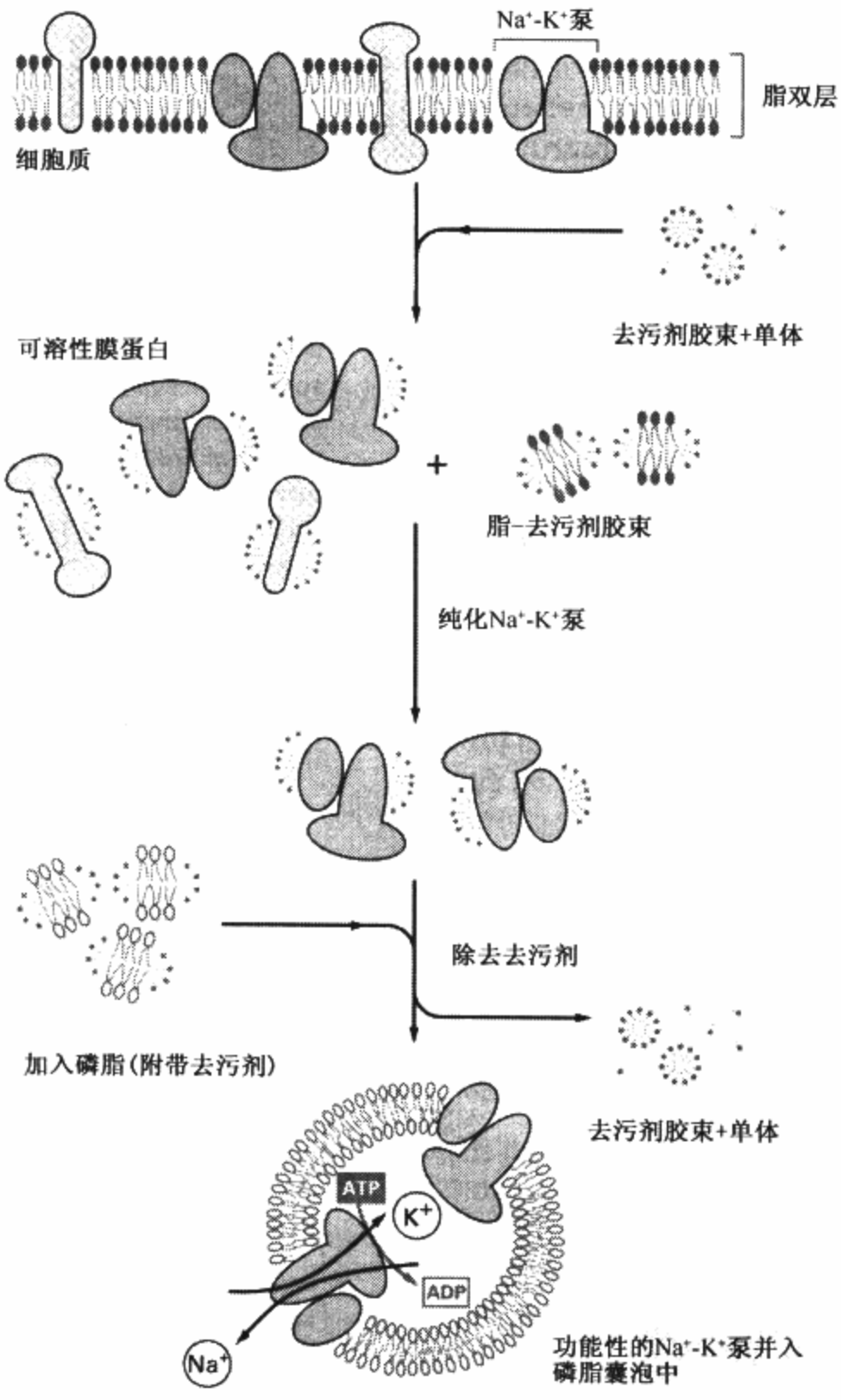


图 10-26　中性去污剂用于溶解、纯化和重新组成功能性膜蛋白系统

在这个例子中，功能性的 Na^+-K^+ 泵分子被纯化并被并入磷脂囊泡中。Na^+-K^+ 泵是一种存在于大多数动物细胞的质膜中的离子泵；它使用 ATP 水解的能量将 Na^+ 泵出，将 K^+ 泵入细胞，讨论见第 11 章

用血影细胞来研究胞质内侧的质膜蛋白

关于人红细胞（图 10-27）质膜的了解多于关于其他任何真核细胞膜的了解。关于这个有好几个原因。例如，从血库中可以获得大量的不含有其他细胞类型的红细胞（也叫做红血球）。因为红细胞不具有细胞核和其他细胞器，质膜是细胞唯一的膜，它在分离后将不会混杂有细胞内膜（因此避免了从其他真核细胞中准备质膜的一系列问题，在那些细胞中质膜在细胞膜的含量要少于 5%）。

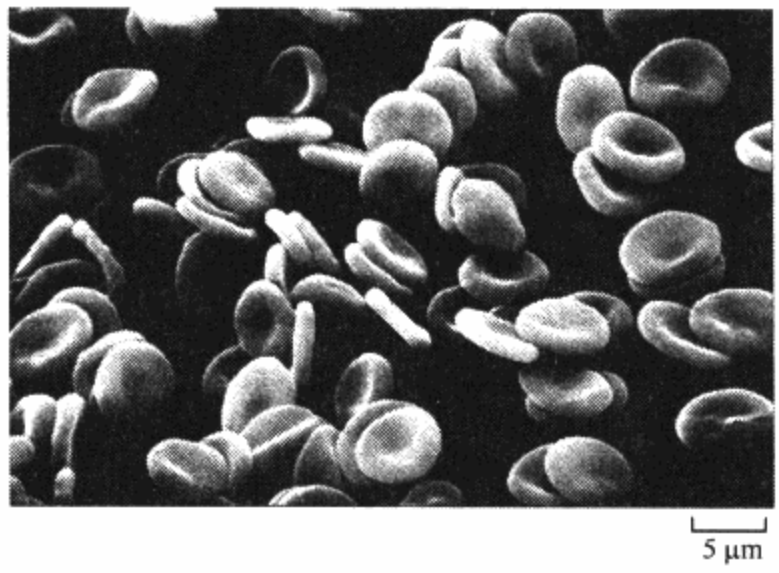


图 10-27　人红细胞的扫描电子显微镜照片

细胞有双凹形状，不含细胞核和其他细胞器

通过将细胞放到一个盐浓度低于细胞内部的溶液中，很容易就能准备空的红细胞膜，或者叫血影细胞。然后水流进红细胞，因此红细胞膨胀并破裂（细胞溶解），将血红蛋白和其他细胞质中的蛋白质释放出来。当这种膜还是漏的时候可以用于研究（那些能够和膜两面结合的试剂被用在这种情况中），这种膜也能重新密封，这样那些水溶性的试剂就到不了内表面。此外，也能利用血影细胞准备反转封装（sealed inside-out）的囊泡（图 10-28），因此能够分别研究膜的内侧（胞质一侧）和外侧。使用密封的和未密封的血影细胞首次证明了有些膜蛋白是横跨脂双层的（接下来会提到）以及双分子层的两半中的脂分子组成是不同的。像这样许多最初在红细胞膜中证明的基本原理，后来发现也适用于真核细胞和细菌的许多膜上。

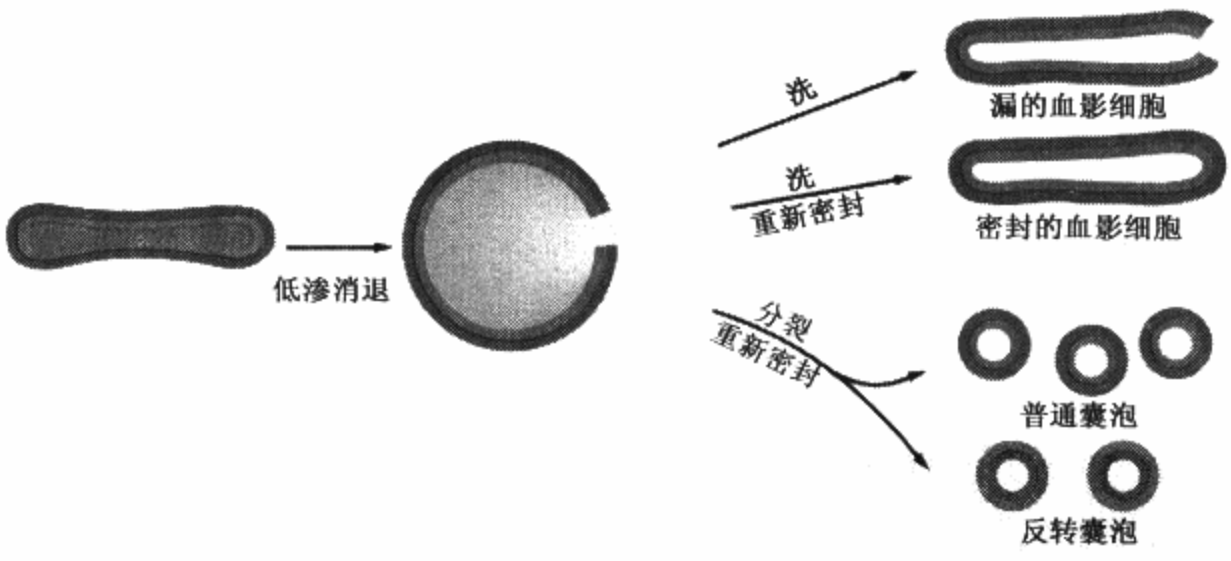


图 10-28　密封和未密封的血影细胞的准备以及普通囊泡（right-side-out vesicle）和反转囊泡（inside-out vesicle）的准备

膜蛋白的“面”能够通过好几种方法测定。一种方法是使用一类共价标记试剂（譬如放射性或荧光的标记物），这种试剂是水溶性的，因此不能透过脂双层；这种试剂只能共价结合到蛋白质暴露在膜外面的那部分。然后用去污剂溶解膜，通过 SDS-PAGE 分离这种蛋白质。这种标记的蛋白质能通过放射性（凝胶中放射自显影）或荧光（将凝胶暴露在紫外光下）来检测。通过使用这种向量的标记，检测凝胶中蛋白质条带，能够测出特定的蛋白质如何在膜上确定方向。例如，如果它从胞内侧和胞外侧都被标记上，那么它一定是一个跨膜蛋白。一

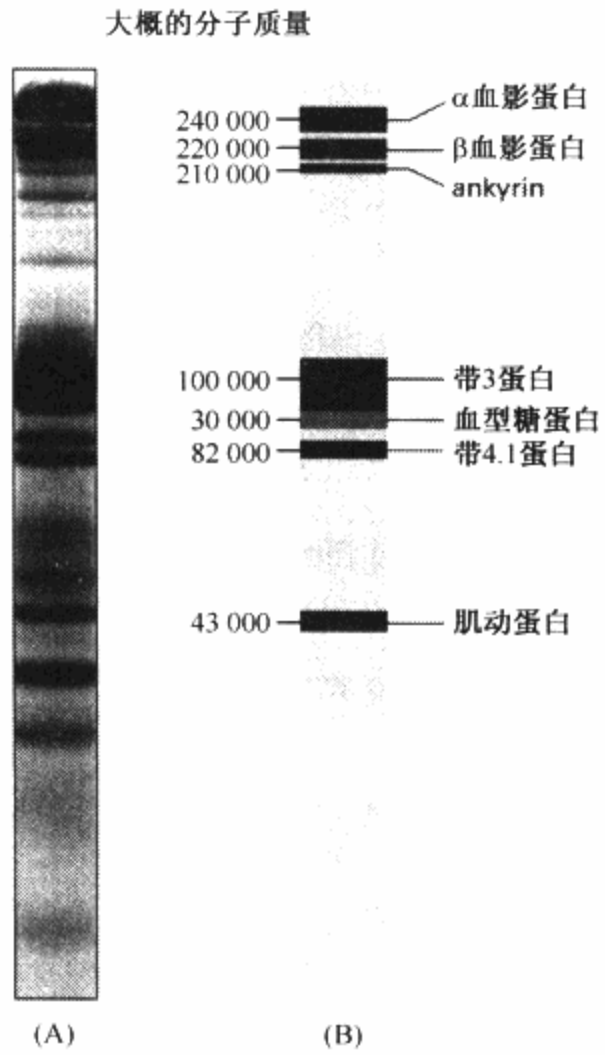


图 10-29　人红细胞膜蛋白的 SDS-PAGE 图谱

（A）凝胶用考马斯亮蓝染色；（B）凝胶中主要蛋白所在的位置表现在示意图中。血型糖蛋白用红色显示，和带 3 蛋白区别开。凝胶中其他的一些条带在示意图中被略去。血型糖蛋白上的大量带负电的糖基降低了它的迁移率，所以血型糖蛋白跑的几乎和大得多的带 3 蛋白一样慢（A，经 Ted Steck 许可）

个可选用的方法是将胞内侧和胞外侧分别暴露在蛋白水解酶中，这种酶是不能透过膜的：如果一个蛋白质从两个表面都被消化，那它一定是跨膜蛋白。此外，只能和蛋白质某部分结合的标记抗体，也能用于检测跨膜蛋白暴露在膜外的两侧的部分是不是抗体所能结合的那部分。

使用 SDS-PAGE 研究人红血球的膜蛋白时，大约检测到 15 种主要的蛋白质条带，分子质量从 15～250kDa 变化，这些蛋白质中有三种——血影蛋白、血型糖蛋白和带 3 蛋白，占了所有膜蛋白重量的 60%以上（图 10-29）。这些蛋白质每种都以一个不同的方式排列在膜中。所以我们将以这些蛋白质为例来说明蛋白质和膜结合的三种主要的方式，这不仅仅适用于红血球，而且也适用于其他细胞。

血影蛋白是一个细胞骨架蛋白，通过非共价作用和红细胞膜的胞质面相连

大部分与人体红血球细胞膜结合的蛋白质分子都是与脂双层胞质侧结合的外周膜蛋白。这类蛋白质中含量最丰富的是**血影蛋白**，这是一种又长又细的柔软的杆状的 100nm 长的蛋白质，它大概构成了膜结合蛋白的质量的 25%（一个细胞中大概有 2.5×10^5 个血影蛋白）。它是蛋白质网络（细胞骨架）中的主要组成部分，它作为红细胞膜的构成基础，维持着这种细胞膜的两面凹陷的形状和结构的完整（图 10-26）。如果在低离子浓度的溶液内将细胞骨架从血影细胞中分离出来，那么细胞膜就会碎裂成小囊泡。

血影蛋白是两种大型的、结构相似的亚基组成的异聚体（heterodimer）（图 10-30）。这些异聚体头对头自我连接形成了 200nm 长的四聚体。四五个四聚体通过结合肌动蛋白纤维和其他的细胞骨架蛋白质（包括带 4.1 蛋白）的形式又相互连接，而形成了一个“连接复合体”。最终所形成的是一个可以变形的网络状结构，它是构成细胞膜的整个胞质表面的基础。就是这种血影蛋白为基础的细胞骨架使红血球可以在被挤压通过毛细血管的时候经受住它的膜的压力。在血影蛋白上出现基因异常的人和鼠会患有贫血，并且他们的红血球的形状不是两面凹陷而是球形的，而且十分脆弱。血影蛋白越缺乏，贫血症状就越严重。

通过监测放射性标记的血影蛋白在那些已经去除了血影蛋白和其他的外周膜蛋白的红细胞膜的结合，鉴定出血影蛋白细胞骨架附在细胞质膜上主要是依靠蛋白质的作用。这些试验表明血影蛋白的结合取决于一种细胞内附属的大蛋白质，叫**锚蛋白**。锚蛋白既与血影蛋白相连又与跨膜蛋白带 3 蛋白的胞质区相连（图 10-31）。通过将带 3 蛋白与血影蛋白相连，锚蛋白就将血影蛋白网络与细胞膜相连了。这也大大减少了脂双层中绑定的带 3 蛋白的扩散率。以血影蛋白为基础的细胞骨架也通过另一种机制与细胞膜相连，这种机制依靠上文提到的带

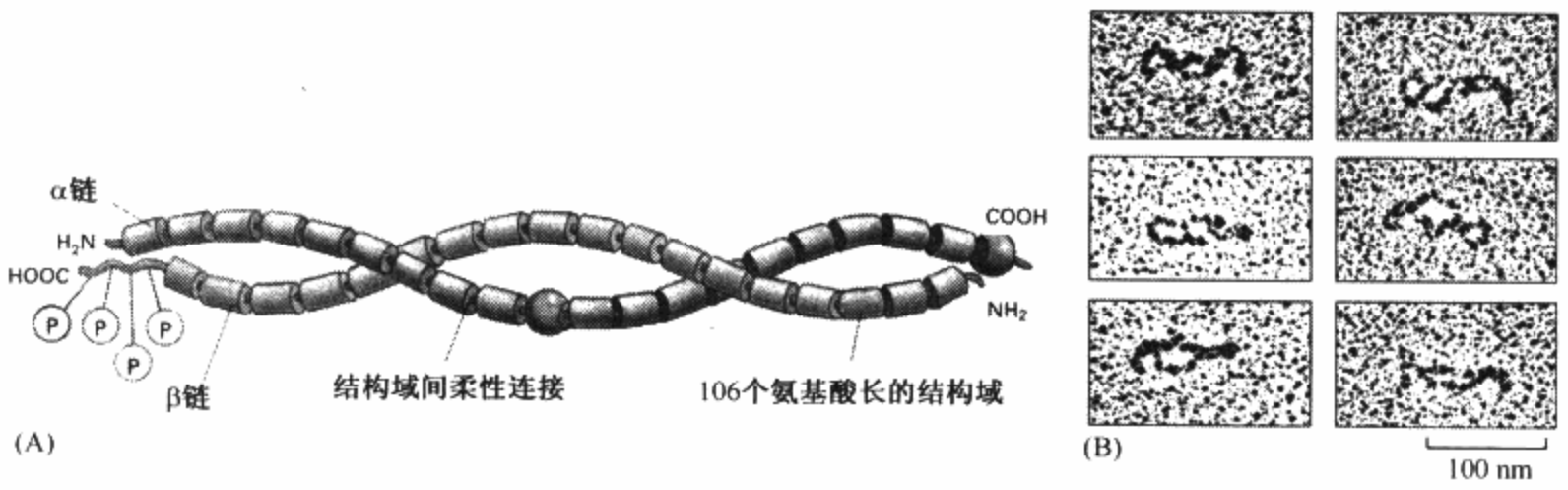


图 10-30　人红细胞中的血影蛋白

蛋白质分别用示意图（A）和电子显微镜照片（B）显示。每个血影蛋白二聚体是由两条方向平行、松散缠绕的柔软的多肽链组成，两条链分别称为α链和β链。两条链在多个位置通过非共价键结合。两个二聚体可以通过磷酸化的头部末端形成四聚体。α链和β链大部分都是由 106 个氨基酸长的重复区域组成。在显微照片中，血影蛋白已经被用铂投影

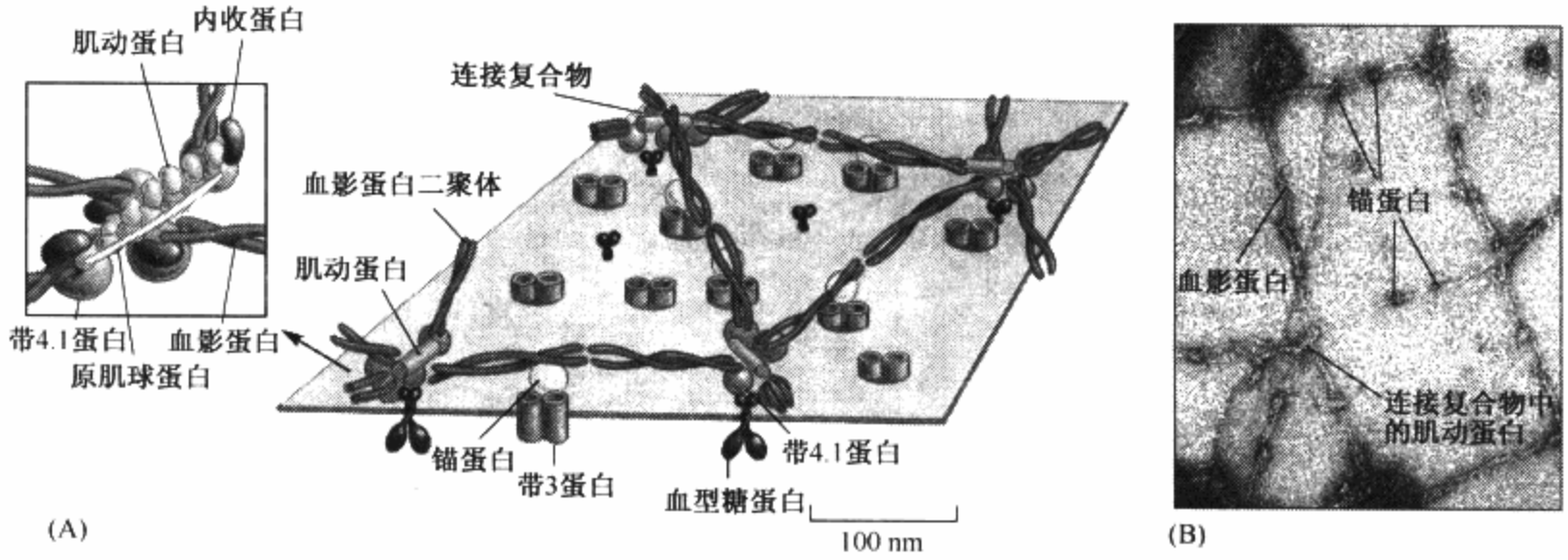


图 10-31　在人红细胞膜胞质侧的基于血影蛋白的细胞骨架

用示意图（A）和电子显微照片（B）显示结构。显示在示意图中的排列主要是从那些对纯化后的蛋白质在体外的相互作用的研究中推出。血影蛋白二聚体通过连接复合物（在左边的图框中有放大图）连接成网状结构。连接复合物由短的肌动蛋白纤维（包括 13 个肌动蛋白单体）、带 4.1 蛋白、内收蛋白和一个原肌球蛋白组成。电子显微照片显示固定并负染后的红细胞膜胞质侧的细胞骨架。为了能更好地观察结构细节，血影蛋白网状结构被特意拉开。普通细胞中，网状结构更拥挤，只占现在面积的 1/10［（B）由 T. Byers and D. Branton 提供，*Proc. Natl. Acad. Sci. USA* 82：6153～6157，1985. © National Academy of Sciences］

4.1 蛋白。这种蛋白质连接血影蛋白与肌动蛋白，但同时也连接带 3 蛋白的胞质区以及血型糖蛋白，血型糖蛋白是红细胞中另外一种主要的跨膜蛋白。

一个相似的，但更精细也更复杂的细胞骨架网络存在于我们体内大部分细胞的质膜底下。这个网络，构成了胞质溶液的皮层，其中富含肌动蛋白丝，这种细丝被认为能通过多种方式连接质膜。我们将在第 16 章讨论那些存在于真核细胞的皮层中，与血影蛋白、锚蛋白和带 4.1 蛋白在结构上同源的蛋白质以及这些蛋白质和质膜的相互作用。

血型糖蛋白用一个单个的α螺旋穿过红细胞的脂双层

血型糖蛋白是暴露在人体红细胞外表面的两种主要蛋白质之一，也是第一种被测定出氨基酸全序列的膜蛋白。如图 10-22 中所显示的跨膜蛋白模型那样，血型糖蛋白是一种单次跨膜的

小型糖蛋白（有 131 种氨基酸），它的大部分都在细胞膜的外表面，这里也是亲水性 N 端所在的地方。蛋白质的这个部分携带了所有的糖基（在 16 个单独的低聚糖侧链中大约有 100 个糖残基），这些糖基占该分子质量的 60%。事实上，红细胞表面糖基的绝大部分（超过 90%的唾液酸和表面大部分的负电荷）是由血型糖蛋白分子携带的。而血型糖蛋白的亲水 C 端暴露在胞质中，而该蛋白质穿过脂双层的是 23 个氨基酸长的 α 螺旋疏水片段［图 10-20（A）］。

尽管一个细胞里有将近 100 万个血型糖蛋白分子，但它们的功能仍不为人所知。可是事实上，如果人体的红血球细胞缺少这种分子的主要亚类，人体也可以非常健康。尽管血型糖蛋白只在红细胞里才能找到，可是它的结构却是一种普通类型的膜蛋白的结构，这种膜蛋白是由一个单个的 α 螺旋穿过脂双层的。例如，许多细胞表面受体都属于这一类。

血型糖蛋白通常作为一个同型二聚体而存在。它的两个相似的链主要是通过非共价相互作用在跨膜的 α 螺旋中连接的。这样，膜蛋白的跨膜片段总是多于一个疏水性的锚：疏水氨基酸的序列可能包含了介导蛋白质间的相互作用的信息。同样地，在多次跨膜蛋白中的单个跨膜片段在折叠后的蛋白质结构中占据着特定的位置，那能够由周围的 α 螺旋之间的相互作用所确定。通常，多次跨膜蛋白的跨膜片段在胞质或非细胞质中的之间通过的茎环结构的多肽链相连，这种多肽链会被蛋白酶切断，产生的片段停留在一起的话还能正常执行功能。有些情况下，碎片可以传递到细胞中，它们自己又可以适当地组合形成一个具有功能的蛋白质（图 10-32）。

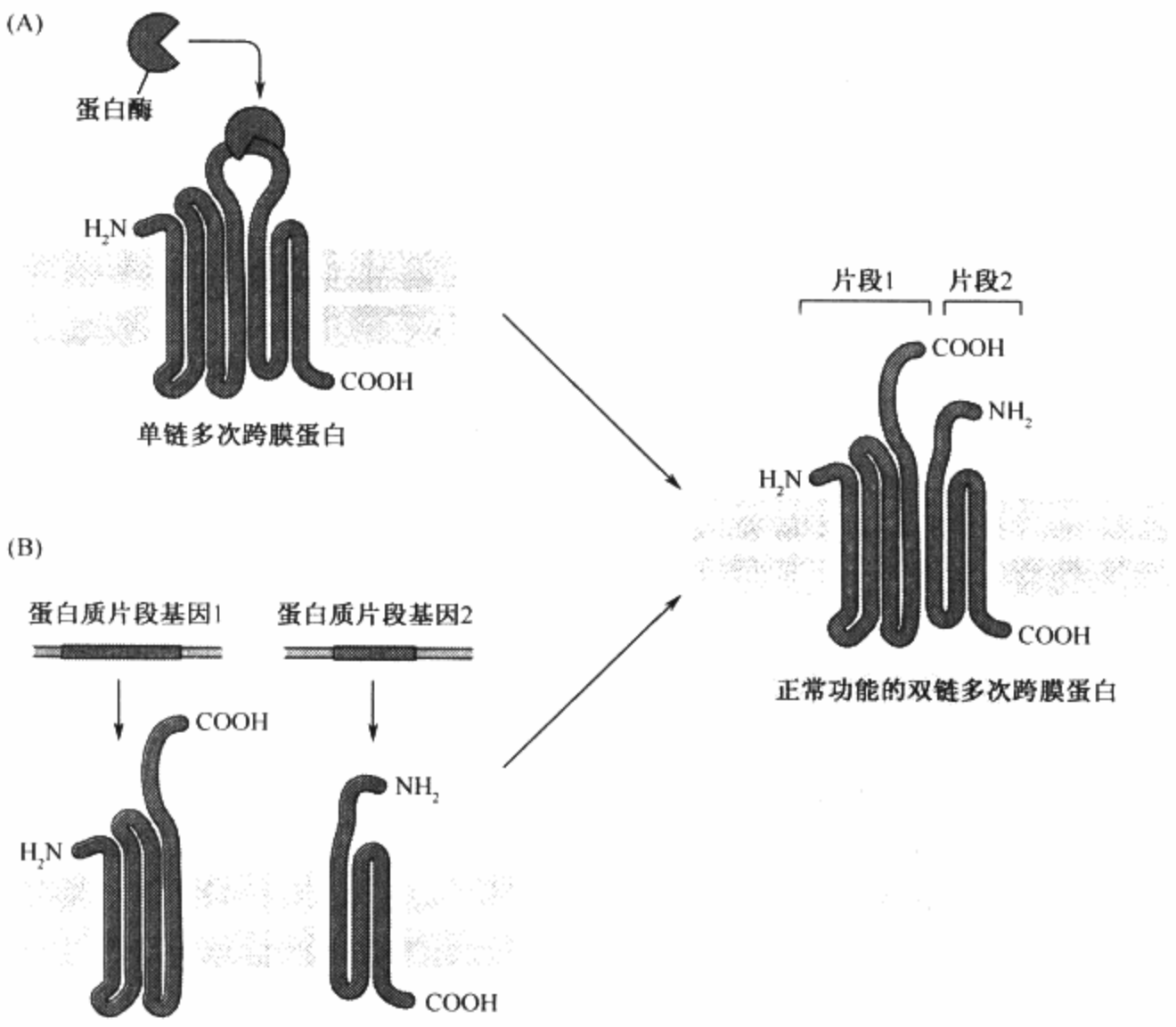


图 10-32　单链多次跨膜蛋白向双链多次跨膜蛋白的转变

（A）蛋白质的一个茎环被水解切开成两个片段，两个片段保持在一起，仍能正常执行功能；（B）从分开的基因中表达了和（A）中相同的两个片段，得到了一个正常执行功能的相似蛋白质

红细胞中的带 3 蛋白是一个促进阴离子转运的多次跨膜蛋白

与血型糖蛋白不同，**带 3 蛋白**被认为在红细胞中具有重要作用。它得名于在 SDS-聚丙烯酰胺凝胶中和其他膜蛋白的相对位置（图 10-29）。与血型糖蛋白相同的是，带 3 蛋白是一种跨膜蛋白，但它是一种多次跨膜蛋白，它用一种高度折叠的构象横跨细胞膜。一般认为它的多肽链（大概 930 个氨基酸）横穿过双分子层 12 次。

红细胞的主要功能是从肺部运输氧气到各个组织，并且将二氧化碳从各个组织输送回肺部。带 3 蛋白在第二个功能中的作用至关重要。因为二氧化碳难溶于水，带 3 蛋白在血浆中的运输是以碳酸氢盐的形式来传输的。红细胞中有一种酶可以通过催化 $H_2O + CO_2 \rightleftharpoons HCO_3^- + H^+$ 的反应来进行二氧化碳和水的组合和分解。带 3 蛋白起着一个阴离子运送者的作用，它通过交换氯离子让碳酸氢根离子穿过细胞膜。通过提高了红细胞膜对碳酸氢根离子的透过性，这个输送者能够增加血液输送给肺部的二氧化碳的含量。

通过**冷冻断裂电子显微技术**（freeze-fracture electron microscopy），可以看到带 3 蛋白是明显的膜内粒子。在这个实验过程中，细胞在液氮中冷冻，冻成的冰块用刀切开，这一裂开的表面倾向于沿脂双层的疏水中心裂开的，这样就把脂类双分子分成了两个单层（图 10-33）。暴露的断面用铂来投影，再用电子显微镜观察这个铂层，用这种方式观察可以看到，人体的红血球细胞膜上镶嵌着许多膜内粒子，它们的直径差不多 7.5nm，并且随机分布（图 10-34）。这些粒子被认为主要是带 3 蛋白。当合成脂双层通过重组过程带上纯化的带 3 蛋白时，双分子层断裂后，就可以通过显微镜看到典型的 7.5nm 长的膜内粒子。图 10-35 显示了为什么冰冻断裂电子显微技术可以看到红细胞膜上的带 3 蛋白，却看不到血型糖蛋白。

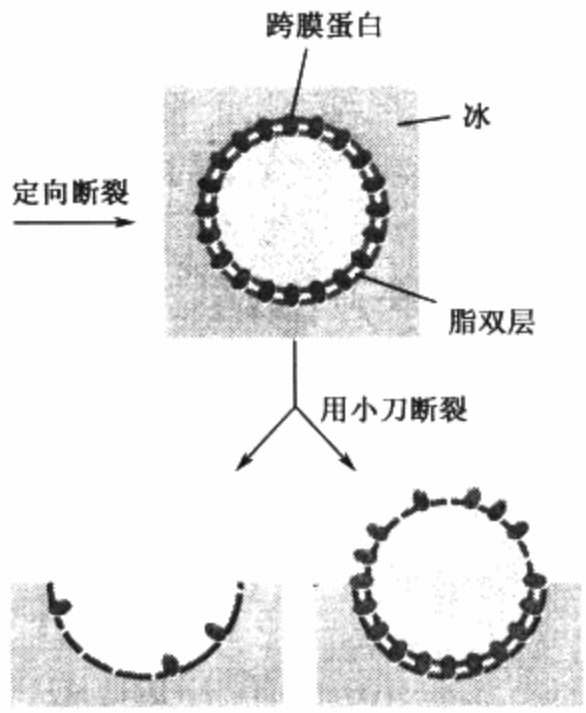


图 10-33　冰冻断裂电子显微成像
这幅图显示了这项技术如何既能显示双分子层在胞质层的疏水部内半层（P face）的图像，又能显示双分子层外层的疏水部外半层（E face）的图像。在断裂操作后，用铂或碳对断裂面投影，有机物随后被消化掉，产生的铂复制层通过电子显微镜检测（图 9-32）

为了将疏水小分子转运穿过细胞膜，一个膜转运蛋白必须穿透脂双层的疏水屏障，并为亲水分子的穿过提供一通道。正如之前讨论过的成孔 β 桶状蛋白一样，多次跨膜蛋白的分子结构也很适合完成这一任务。在很多多次跨膜蛋白质里，一些跨膜的 α 螺旋既包括疏水的也包括亲水的氨基酸侧链，疏水侧链所在的一侧，暴露给在细胞膜的脂分子层。亲水侧链聚集在另一侧，在那里它们构成了部分的亲水管孔的衬里，而那些亲水管孔是由一些这样的两性 α 螺旋在脂双层的疏水中心并排成环而形成的。

在第 11 章中，我们将考虑为什么多次跨膜蛋白会被认为介导了亲水小分子特定的跨膜转运。但如果我们想详细了解膜转蛋白实际上是如何工作的我们就需要有关双分子层的三维结构的精确信息。其中一种像这样的细节已知的质膜蛋白叫细菌视紫红质，在一些古细菌的质膜中它是作为一种光活化的质子泵。细菌视紫红质的结构与其他的许多膜蛋白相似，接下来我们简单地讨论它。

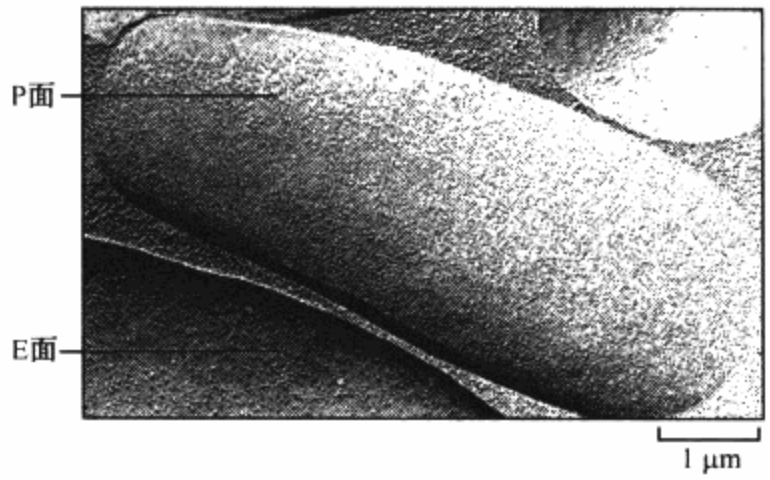


图 10-34　人红细胞的冰冻断裂电子显微照片

需要注意的是膜内颗粒的密度在胞质面（P）的比外层膜（E）的要高

（经 L. Engstrom 和 D. Branton 许可）

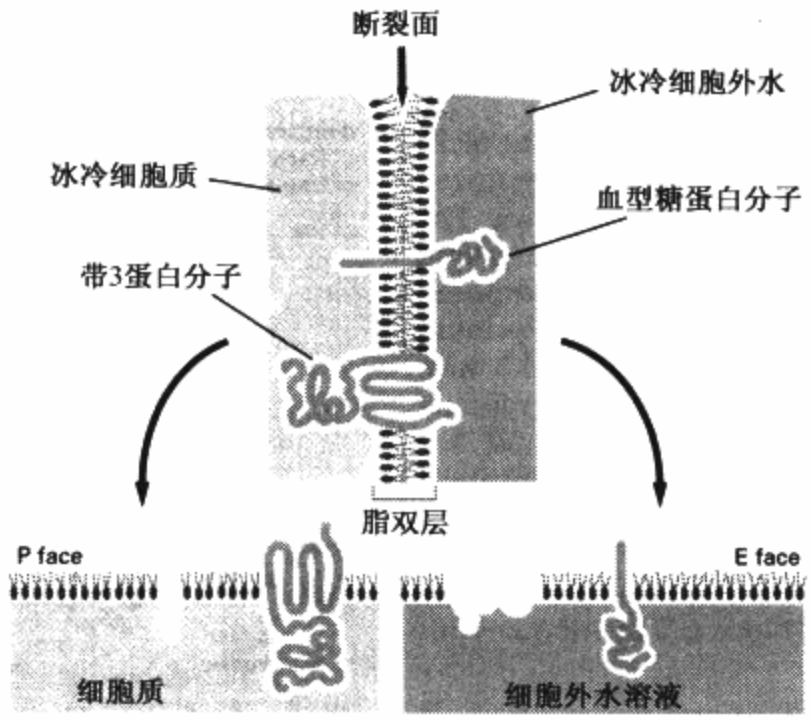


图 10-35　当冰冻断裂时，在人血细胞膜中的带 3 蛋白和血型糖蛋白可能的命运

人血细胞膜中的 band 3 和血型糖蛋白分子在冰冻断裂时的可能命运。当脂双层被裂开时，每个跨膜蛋白分子的膜内部分和膜外部分从相关的脂单层脱离，而蛋白质分子则倾向于滞留在包含了大部分蛋白的脂单层中。由于这一原因，band 3 分子通常滞留于内小页断裂面，即 P 面；且该分子在断裂面上形成的团块足够大，因此，band 3 分子是易于识别的膜内颗粒。血型糖蛋白分子通常滞留于外小页断裂面，即 E 面，但一般认为该分子的胞质溶胶尾不够大，因而在断裂面上不可见。

细菌视紫红质是一个质子泵，通过 7 个 α 螺旋穿过脂双层

古细菌嗜盐菌的“紫膜”是质膜里特殊的一小片，但它却包含着一类蛋白质分子——**细菌视紫红质**（图 10-36）。每个细菌视紫红质分子包含一个单个的光吸收基团，或者叫做生色团（叫视网醛），这一物质让蛋白质呈现紫色。视网醛是一种醛类形式的维生素 A，它与在脊椎动物眼睛中的感光细胞中的视紫质中发现的生色团很相似（讨论见第 15 章）。视网醛与细菌视紫红质蛋白中的赖氨酸侧链共价相连。当被光线中的单光子激活时，激发态的生色团会改变形状并导致蛋白质中一系列小的结构变化，这些变化

图 10-36　嗜盐杆菌的示意图显示了包含有细菌视紫红质分子的紫膜斑块

这些古细菌生活在盐水中，暴露在阳光下，已经进化出一系列的光活化蛋白，包括细菌视紫红质这种在脂膜中的光活化质子泵

会帮助把氢离子从细胞里转运到细胞外（图 10-37）。在强光下，每一个细菌视紫红质分子可以在一秒内泵出几百个质子。这些光驱动的质子转运建立了一个跨膜的氢离子梯度，它反过来又会促进细胞质膜里另一个蛋白质的 ATP 的生产，以及其他能够利用这种氢离子梯度里所储藏的能量的过程。因此，细菌视紫红质可以说是给细菌细胞提供能量的太阳能传感器的一部分。

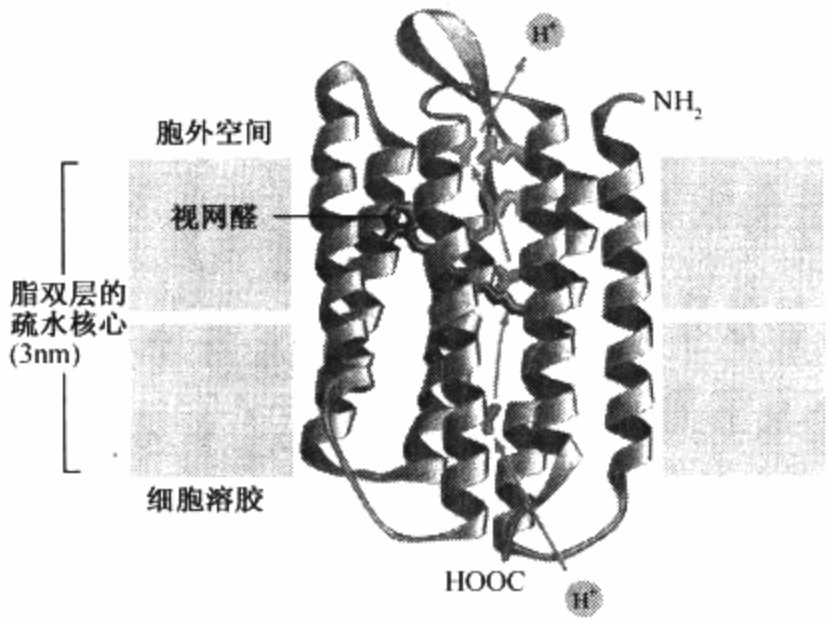


图 10-37　细菌视紫红质分子的三维结构

多肽链以 α 螺旋的形式穿过脂双层 7 次。图中显示了视黄醛生色团的位置（紫色）和光活化的泵循环中质子可能的通道。最开始也是最关键的一步是在生色团吸收一个光子后，一个质子被从生色团传递到 85 位的天门冬氨酸的侧链上（红色，位于生色团旁边）。随后利用那些跨膜排列成一个通道的亲水性氨基酸侧链，这些质子传递者完成泵循环，使得酶恢复到起始状态。颜色代码：谷氨酸（橙色），天门冬氨酸（红色），精氨酸（蓝色）（改编自 H. Luecke et al.，*Science* 286：255～260，1999）

要通过分子细节了解多次跨膜蛋白的功能，就必须精确地定位它的每个原子，这一般需要对排列秩序的蛋白质三维晶体进行 X 射线衍射研究。但因为它们的两性分子的性质，这些蛋白质只在去污剂溶液中可溶，而且很难结晶。然而，紫膜中的大量的细菌视紫红质分子是按照平面二维方式结晶的。通过使用电子显微镜与电子衍射分析相结合，常规的配置已经可以高分辨率（3Å）地测定蛋白质的三维结构以及细菌视紫红质在膜中的定向。这一过程被称为**电子晶体学**，它与通过 X 射线衍射分析研究可溶蛋白质的三维晶体结构相类似。通过电子晶体学获得的这个结构随后被验证，而且比 X 射线晶体学有更高的分辨率。这些研究表明了每个细菌视紫红质分子都折叠成 7 个紧密排列的 α 螺旋，每个螺旋都包含 25 个氨基酸，它们穿入脂双层的角度都有所不同。通过在较低温度下冷冻蛋白质结晶的方式，了解蛋白质在泵作用循环中的中间构象已经成为可能。

细菌视紫红质是膜蛋白大家庭中的一员，它同膜蛋白家族中的许多蛋白质有相似的结构却有不同的功能。举例来说脊椎动物视网膜中的杆状细胞中的视紫质，以及许多能结合胞外信号分子的细胞表面受体蛋白，都是在 7 个跨膜的 α 螺旋的结构上组成的。这些蛋白质的功能与其说是信号转运不如说是信号传感器：每一个蛋白质通过激活细胞内另一个蛋白质的方式来对胞外信号应答，这种激活可以在胞质里产生化学信号，正如我们在第 15 章里要讨论的一样。尽管细菌视紫红质的信号受体与哺乳动物的信号受体的结构十分相似，它们在序列上却有很大不同，因此它们可能分属于一个古老的蛋白质家族里两个完全不同的进化分支。

膜蛋白通常形成大复合物来行使功能

一些膜蛋白作为多组分的复合物的部分来行使功能。其中一些已经通过 X 射线晶体学得以研究。其中一个是细菌的光合作用中心，这是第一个被结晶和 X 射线衍射分析的蛋白质复

合物。这一分析的结果对于膜蛋白来说有普遍重要性，因为这是第一次显示了多个多肽如何在膜中相互联合形成复杂的蛋白质机器（图 10-38）。在第 14 章我们要讨论像这样的光合复合物如何捕获光能并将 H^+ 泵出膜外，同时也要讨论那些比光合作用中心还要大的膜蛋白复合物的结构和功能。膜蛋白通常排在大的复合物中，不仅仅用于捕获多种形式的能量也能用于将胞外信号转换成胞内信号（讨论见第 15 章）。

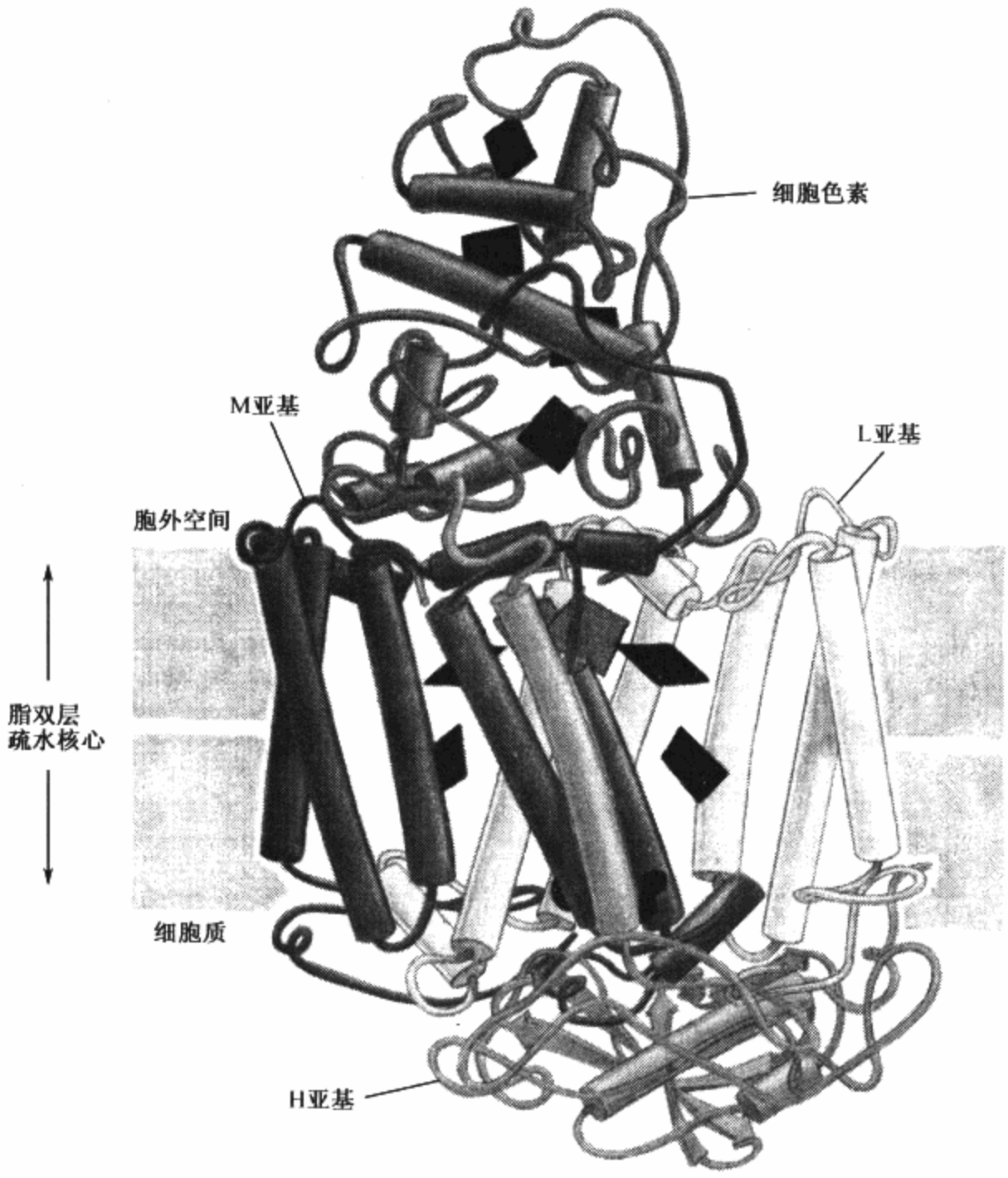


图 10-38　细菌 *Rhodopseudomonas viridis* 的光合反应中心的三维结构

这个结构通过 X 射线衍射分析这种跨膜蛋白复合物的晶体获得。复合物包括 4 个亚基：L 亚基、M 亚基、H 亚基和 1 个细胞色素。L 和 M 亚基形成反应中心的核心，每一个都包括 5 个跨膜的 α 螺旋。图中黑色部分显示的是不同的电子传递辅酶位置所在。需要注意的是辅酶排列在这些螺旋之间的空间中（改编自 J. Richardson，基于 J. Deisenhofer, O. Epp, K. Miki, R. Huber, and H. Michel. *Nature* 318：618～624，1985）

许多膜蛋白在膜平面中扩散

像膜脂分子一样，膜蛋白不能在脂双层中翻转，但是它们能够绕一和膜平面垂直的轴旋转（转动扩散）。此外，很多膜蛋白能够在膜中侧向移动（侧向扩散）。一些质膜蛋白可以膜平面中移动的第一个直接的证据是通过一个实验来证明的，在这个实验中将小鼠细胞和人细胞人工融合，产生杂交细胞（异核体）。两种不同标记的抗体被用来选择性地区分小鼠和人质膜蛋白。尽管在新形成的异核体中小鼠和人的蛋白质开始时被限制在它们原来所属的那一半

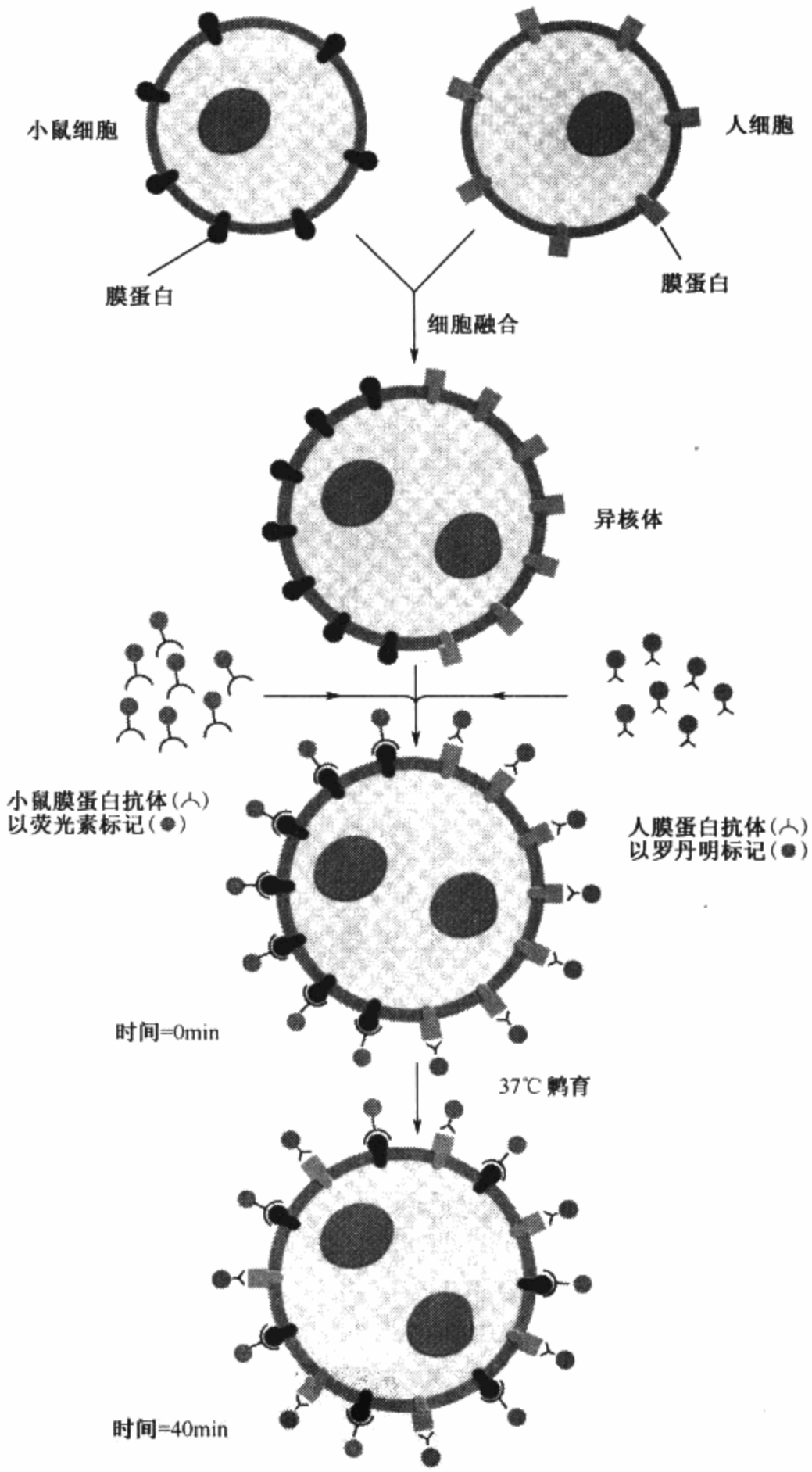


图 10-39　用人鼠杂交细胞证明质膜蛋白的混合的一个实验

在鼠和人的蛋白质开始时分别被限制在新形成的异核体脂膜中它们原来所在的那一半内。但是随着时间它们开始混合。用来指示蛋白质的这两种抗体可以在荧光显微镜中区分，因为荧光素是绿色而罗丹明是红色的（基于 L. D. Frye and M. Edidin 的描述，J. Cell. Sci. 7：319～335，1970. © The Company of Biologists）

中，但大约 30min 之后，这两类蛋白质扩散并混合到整个细胞的表面（图 10-39）。

膜蛋白侧向扩散的速率可以通过光漂白后荧光恢复（FRAP）的技术来测量。这方法通常是用特殊的荧光基团来标记要研究的膜蛋白。这种方法的实现可以通过用抗体之类的荧光配体和

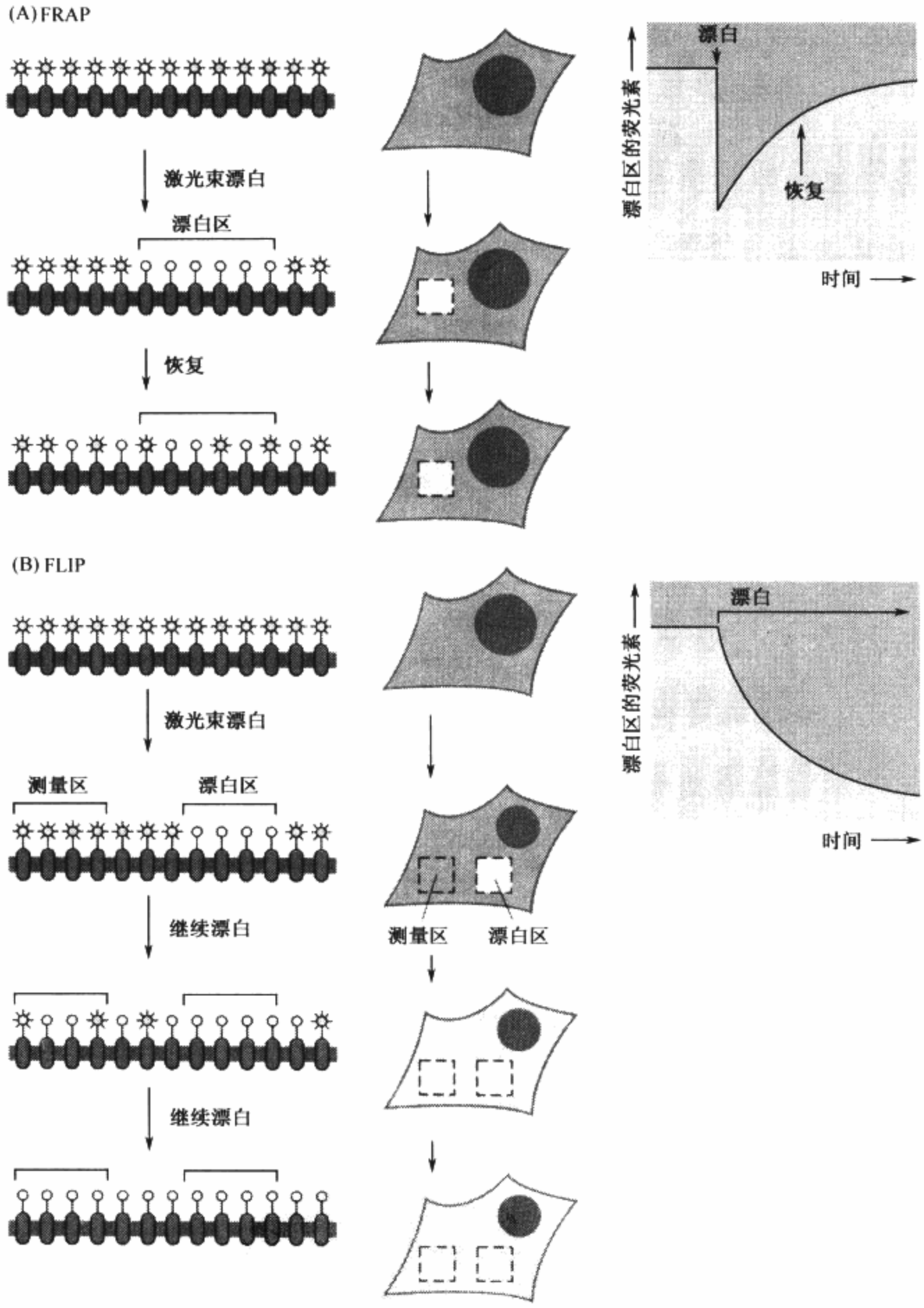


图 10-40　使用光漂白技术测量膜蛋白侧向扩散的速率

一个特定的膜内蛋白可以用荧光抗体标记（正如图中所给出的），或者可以表达一个带有绿色荧光蛋白（GFP）的融合蛋白。（A）用 FRAP 技术，使用一束激光将一小块区域的荧光分子漂白，随着被漂白的分子扩散出这片区域以及未漂白的分子扩散进入这片区域，荧光强度恢复。扩散常数可以从恢复率的图中计算出来：膜蛋白的扩散常数越大，恢复越快。（B）用 FLIP 技术，连续照射膜的一个区域，荧光测量在另一个区域。随着荧光蛋白扩散出去而漂白了的分子扩散进来，在第二个区域中的荧光逐渐衰弱；只要所有的荧光蛋白分子都是可移动的，没有锚定在细胞骨架或胞外基质上的话，最终所有的荧光蛋白分子都会被漂白

蛋白质的结合或利用重组 DNA 技术表达和 GFP 融合后的蛋白质（讨论见第 9 章）。然后用激光将一小片区域中的荧光基团漂白，测量临近区域中携带有未漂白的配体和 GFP 的蛋白质扩散进入漂白后的区域所需的时间［图 10-40（A）］。一个互补的技术是光漂白荧光缺失技术（FLIP）。这种方法中，用一束激光连续照射一小块区域来漂白所有进入该区域的荧光分子，在那附近逐渐耗尽周围膜中所有的荧光标记分子［图 10-40（B）］。通过这种测量人们可以计算那些标记的特定细胞表面蛋白的扩散系数。对于不同细胞中的不同膜蛋白扩散系数的值有很大差异，因为和其他蛋白质的相互作用会不同程度地阻止蛋白质的扩散。对那些被最低限度阻止的蛋白质的测量显示细胞膜具有和橄榄油类似的黏度。

细胞可以将蛋白质和脂分子限定在细胞膜中特定区域

对生物膜是二维流体的认识是在理解膜结构和功能中的主要进展。不过，对于那幅蛋白质自由地漂浮在脂分子的海洋中的生物膜图像，人们已经清楚地知道这种认识过分单纯化了。许多细胞有很多将膜蛋白限定在连续脂双层的特定区域中的方法。在上皮细胞中，如那些排列在肠道或肾小管中的细胞，某些质膜酶和转运蛋白被限制在细胞顶端的表面上，而其他则被限制在底面或侧面（图 10-41）。这种膜蛋白不对称的分布通常是上皮细胞功能的基础，我们在第 19 章会讨论这一点。这两个膜区域的脂类成分也是不同的，这证明了上皮细胞既能防止蛋白质分子在这些区域间的扩散也能防止脂分子的扩散。不过，采用标记了的脂分子的实验显示了只有那些在膜外部单层的脂分子才受到这种方式的限制。蛋白质和脂分子的这种分离被认为至少部分是由一种特殊类型的细胞间连接（紧密连接，讨论见第 19 章）所设定的屏障来维持的。很明显，形成这些细胞间连接的膜蛋白是不允许在相互作用的膜中侧向扩散的。

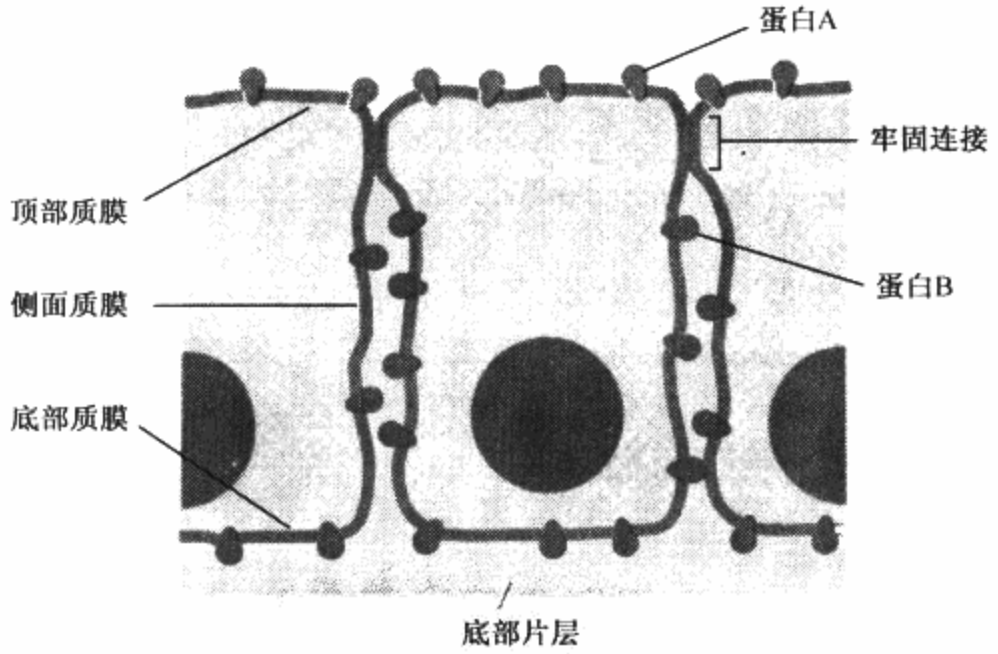


图 10-41　一个质膜蛋白是如何被限制在一个特定的膜区域中

在这个上皮细胞的示意图中，蛋白 A（在顶部的膜中）和蛋白 B（在底部和侧面的膜中）能够在它们所处的区域中侧向扩散，但是不能进入其他区域，至少有部分是被特定的细胞连接限制，这种连接叫牢固连接。质膜的外膜（非胞质面）脂分子单层在这样的两个区域间是不能扩散的，而内膜（胞质面）单层的脂分子却能扩散（未显示）

一个细胞也能不用细胞间连接建立一个膜的区域。例如，哺乳动物的精子是一个由连续膜包被、由几个结构和功能截然不同的区域组成的单个细胞。当采用多种抗体（每种针对一种特定的细胞表面分子）通过免疫荧光显微技术检测一个精子细胞时，发现质膜至少由 3 个不同的区域组成（图 10-42）。一些质膜分子能在它们所被限制的那个区域中自由扩散。防止这些分子

离开所在区域的那些“栅栏”的分子机制还不清楚。许多细胞也有相似的膜栅栏能够限制膜蛋白在某些膜区域中的扩散。例如，神经细胞质膜包含有一个围绕细胞体和树突的区域和另一个围绕轴突的区域。其中认为在细胞体和轴突结合的部位有一个肌动蛋白纤维环带牢固地和脂膜结合，形成了栅栏部分。

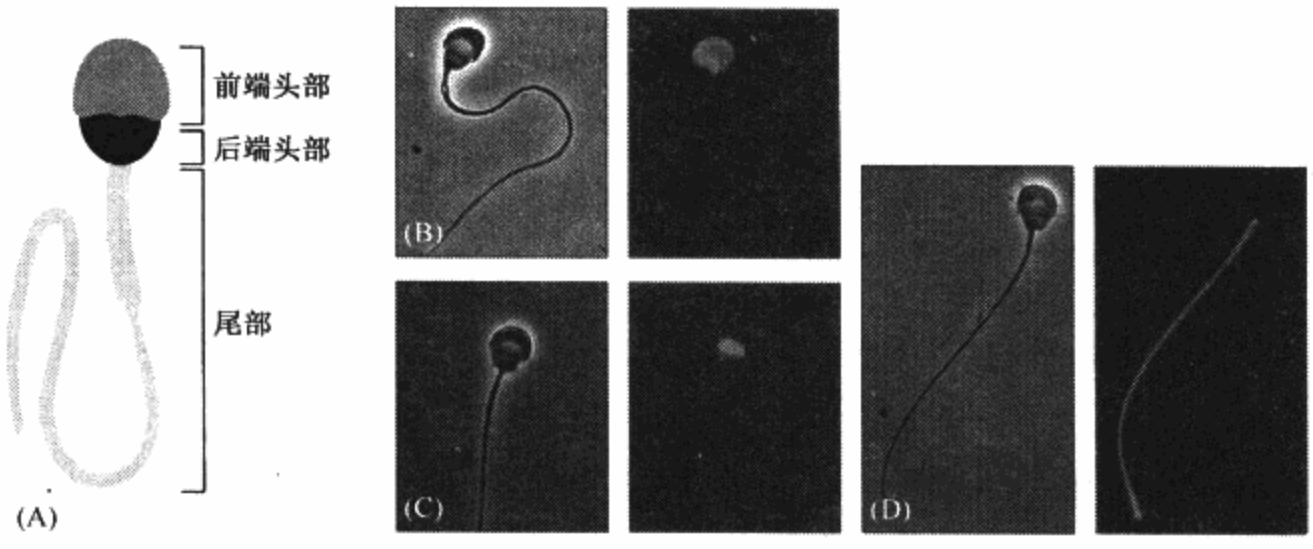


图 10-42　一个豚鼠精细胞质膜的三个区域

（A）一个豚鼠精细胞的基本图像。在三对显微镜照片中，左边的是相差显微镜的照片，右边显示的是相同细胞在细胞表面免疫荧光染色后的照片。不同的单抗选择性标记细胞表面的分子。（B）在头部的前端。（C）在头部的后端。（D）尾巴部分（显微图经 Selena Carroll and Diana Myles 授权）

在所有的这些例子中，蛋白质和脂分子的扩散被限制在连续质膜的特定区域中。已经知道了细胞具有多种方法固定它们的膜蛋白。在盐杆菌的紫色膜中的细菌视紫红质分子聚集成大的二维晶体，其中单个蛋白质分子彼此之间相互固定，这种大聚集体的扩散很慢。一个更常见的限制特定膜蛋白侧向扩散的方法是将蛋白质或脂分子拴在细胞内侧或细胞外侧中的大聚集体上。我们已经知道有些红细胞的膜蛋白被锚定在内部的细胞骨架上。在其他的细胞类型中，质膜蛋白能够被锚定到细胞骨架或细胞外基质上，或者两个都锚定上。4 种大家都知道的固定特定质膜蛋白的方法总结在图 10-43 中。

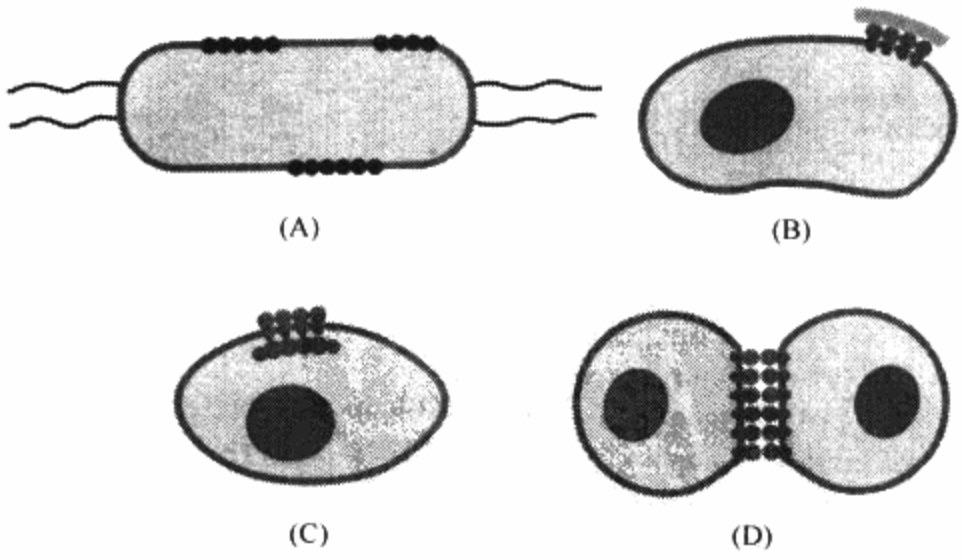


图 10-43　4 种限制特定质膜蛋白侧向移动的方法

（A）蛋白质能自组装成大的聚合体（显示的是细菌视紫红质在嗜盐杆菌的紫膜中）；（B）通过细胞外侧的大分子聚合体的相互作用拴在一起；（C）通过细胞内侧的大分子聚合体的相互作用拴在一起；（D）细胞表面的大分子聚合体能和其他细胞表面的相互作用

细胞表面包被有糖残基

通常质膜蛋白不会毫无修饰地暴露在细胞的表面。它们通常被糖基修饰，这几乎包被在所有真核细胞的表面。这些糖基通常以寡糖链的形式共价结合到膜蛋白（糖蛋白）或脂分子（糖脂）上。它们也会以多聚糖的形式作为膜蛋白聚糖的部分出现。由长的多聚糖链共价连接到蛋白质核心上所形成的蛋白聚糖，主要是在细胞的外侧发现，作为细胞外基质的组成部分（讨论见第 19 章）。但是对于一些蛋白聚糖，它的蛋白质核心可能填充在脂双层中或者通过糖基磷脂酸肌醇（glycosylphosphotidylinositol，GPI）锚定在双分子层上。

术语细胞**外被**或**糖萼**通常用来描述细胞表面糖基富集的地带。这些地带能够被多种染料显色，如钌红（图 10-44），同样因为它对那些被叫做**凝集素**的结合糖基蛋白有亲和力，能够用荧光染料或其他一些可见的标记物来标记。尽管大部分糖基都附在那些固定在脂膜中的分子上，但是糖萼还是包含了那些已经分泌到胞外空间并重新吸附到细胞表面的糖蛋白和蛋白聚糖（图 10-45）。这类被吸附的大分子中许多都是胞外基质的成分，所以质膜的末端和胞外基质的起点

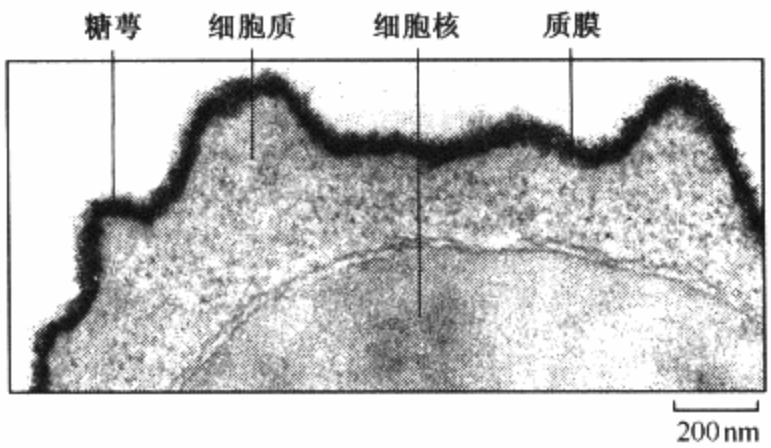


图 10-44　细胞外被或糖萼

这张用钌红染色的淋巴细胞的电子显微镜照片突出了细胞外围厚厚的糖基层

（经 Audrey M. Glauert and G. M. W. Cook. 许可）

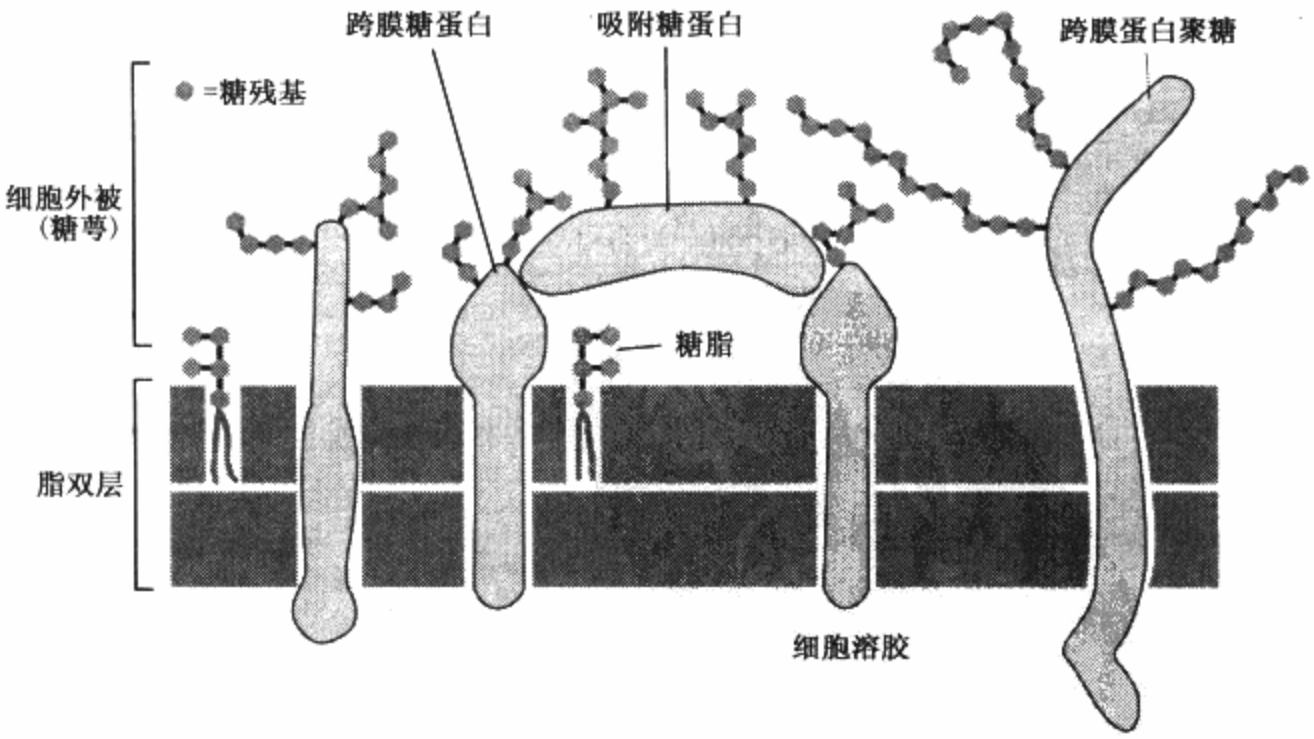


图 10-45　细胞外被（糖萼）简图

细胞外被由糖脂的寡糖侧链和完整的膜糖蛋白以及膜上蛋白聚糖的多聚糖链组成。此外，在许多细胞中糖萼还包括被吸附的糖蛋白和蛋白聚糖（图中未显示）。请注意所有的糖基都在膜的非胞质表面

位置很大程度上是差不多的。细胞外被的一个可能的作用是保护细胞免受机械和化学伤害以及保持细胞和外界物体及其他细胞之间的距离，防止一些不必要的蛋白质之间的相互作用。

糖蛋白和糖脂的寡糖侧链在糖基的排列上多种多样。尽管它们通常包含不超过 15 个的糖基，但它们通常是分岔的，而且糖基还能通过多种共价连接在一起，这和多肽链中通过同样的肽键相连的氨基酸不一样。即便只有 3 个糖基也能形成数以百计不同的三糖。根据这一原理，细胞表面的寡糖的多样性和暴露的位点使得这些寡糖适合担当特异性细胞识别过程的位点。很多年来很少有证据能证明这一猜想的功能。不过最近，能够识别细胞膜表面糖脂和糖蛋白的特定寡糖的凝集素被鉴别出来了。正如我们在第 19 章要讨论的，现在已经知道这些凝集素能够介导许多细胞和细胞之间的黏附过程，包括精子和卵子的相互作用、血液凝集、淋巴细胞再循环以及炎症反应等事件。

小结

尽管脂双层决定了生物膜的基本结构，蛋白质是大多数膜功能产生的原因，蛋白质可以作为特定的受体、酶和转运蛋白等。许多膜蛋白横跨脂双层。这些跨膜蛋白中有些的多肽链通过单个 α 螺旋（单次跨膜蛋白）跨膜；而有些，包括那些跨膜转运离子和其他水溶性小分子的蛋白质，多肽链通过一系列的 α 螺旋或一个形成封闭桶状结构的 β 折叠跨膜多次（多次跨膜蛋白）。其他膜结合蛋白并不跨双分子层，而是附在膜的每一侧。这些蛋白质中许多是通过和跨膜蛋白的非共价作用来结合的，不过也有一些通过和脂分子基团的共价作用来结合。就像脂分子在双分子层中一样，许多蛋白质能够在膜平面中快速扩散。不过，细胞有许多方法来固定特定的膜蛋白以及将膜蛋白和脂分子限定在连续脂双层的特定区域中。

在所有真核细胞的质膜中，多数暴露在细胞表面的蛋白质和一些在外部脂单层的脂分子有共价结合的寡糖链。质膜也含有完整的表面暴露多糖链的蛋白聚糖分子。通常认为糖包被表面的目的是保护细胞表面免受机械和化学伤害。此外，一些寡糖链可以被细胞表面糖基结合蛋白（凝集素）识别，这些蛋白质可以帮助介导细胞间的黏附。

参考文献

一般文献

Bretscher MS (1985) The molecules of the cell membrane. *Sci. Am.* 253(4), 100–109.

Jacobson K, Sheets ED & Simson R (1995) Revisiting the fluid mosaic model of membranes. *Science* 268, 1441–1442.

Lipowsky R & Sackmann E (eds) (1995) The Structure and Dynamics of Membranes. Amsterdam: Elsevier.

Singer SJ & Nicolson GL (1972) The fluid mosaic model of the structure of cell membranes. *Science* 175, 720–731.

The Lipid Bilayer

Bevers EM, Comfurius P, Dekkers DW & Zwaal RF (1999) Lipid translocation across the plasma membrane of mammalian cells. *Biochim. Biophys. Acta* 1439, 317–330.

Chapman D & Benga G (1984) Biomembrane fluidity – studies of model and natural membranes. In Biological Membranes (Chapman D ed) Vol 5, pp 1–56. London: Academic Press.

Devaux PF (1993) Lipid transmembrane asymmetry and flip-flop in biological membranes and in lipid bilayers. *Curr. Opin. Struct. Biol.* 3, 489–494.

Dowhan W (1997) Molecular basis for membrane phospholipid diversity: why are there so many lipids? *Annu. Rev. Biochem.* 66, 199–232.

Hakomori S (1986) Glycosphingolipids. *Sci. Am.* 254(4), 44–53.

Harder T & Simons K (1997) Caveolae, DIGs, and the dynamics of sphingolipid-cholesterol microdomains. *Curr. Opin. Cell Biol.* 9, 534–542.

Ichikawa S & Hirabayashi Y (1998) Glucosylceramide synthase and glycosphingolipid synthesis. *Trends Cell Biol.* 8, 198–202.

Jain MK & White HBD (1977) Long-range order in biomembranes. *Adv. Lipid Res.* 15, 1–60.

Kornberg RD & McConnell HM (1971) Lateral diffusion of phospholipids in a vesicle membrane. *Proc. Natl. Acad. Sci. USA* 68, 2564–2568.

Menon AK (1995) Flippases. *Trends Cell Biol.* 5, 355–360.

Rothman J & Lenard J (1977) Membrane asymmetry. *Science* 195, 743–753.

Simons K & Ikonen E (1997) Functional rafts in cell membranes. *Nature* 387, 569–572.

Tanford C (1980) The Hydrophobic Effect: Formation of Micelles and Biological Membranes. New York: Wiley.

van Meer G (1989) Lipid traffic in animal cells. *Annu. Rev. Cell Biol.* 5, 247–275.

Membrane Proteins

Bennett V & Baines A (2001) Spectrin and ankyrin-based pathways: metazoan inventions for integrating cells into tissues. *Physiol. Rev.* 81, 1353–1392.

Branden C & Tooze J (1999) Introduction to Protein Structure, 2nd edn. New York: Garland.

Bretscher MS & Raff MC (1975) Mammalian plasma membranes. *Nature* 258, 43–49.

Buchanan SK (1999) β-Barrel proteins from bacterial outer membranes: structure, function and refolding. *Curr. Opin. Struct. Biol.* 9, 455–461.

Cross GA (1990) Glycolipid anchoring of plasma membrane proteins. *Annu. Rev. Cell Biol.* 6, 1–39.

Deisenhofer J & Michel H (1991) Structures of bacterial photosynthetic reaction centers. *Annu. Rev. Cell Biol.* 7, 1–23.

Drickamer K & Taylor ME (1993) Biology of animal lectins. *Annu. Rev. Cell Biol.* 9, 237–264.

Driscoll PC & Vuidepot AL (1999) Peripheral membrane proteins: FYVE sticky fingers. *Curr. Biol.* 9, R857–R860.

Edidin M (1992) Patches, posts and fences: proteins and plasma membrane domains. *Trends Cell Biol.* 2, 376–380.

Eisenberg D (1984) Three-dimensional structure of membrane and surface proteins. *Annu. Rev. Biochem.* 53, 595–623.

Frye LD & Edidin M (1970) The rapid intermixing of cell surface antigens after formation of mouse–human heterokaryons. *J. Cell Sci.* 7, 319–335.

Gahmberg CG & Tolvanen M (1996) Why mammalian cell surface proteins are glycoproteins. *Trends Biochem. Sci.* 21, 308–311.

Haupts U, Tittor J & Oesterhelt D (1999) Closing in on bacteriorhodopsin: progress in understanding the molecule. *Annu. Rev. Biophys. Biomol. Struct.* 28, 367–399.

Helenius A & Simons K (1975) Solubilization of membranes by detergents. *Biochim. Biophys. Acta* 415, 29–79.

Henderson R & Unwin PNT (1975) Three-dimensional model of purple membrane obtained by electron microscopy. *Nature* 257, 28–32.

Jacobson K & Vaz WLC (1992) Domains in biological membranes. *Comm. Mol. Cell. Biophys.* 8, 1–114.

Kyte J & Doolittle RF (1982) A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.* 157, 105–132.

Leevers SJ, Vanhaesebroeck B & Waterfield MD (1999) Signalling through phosphoinositide 3-kinases: the lipids take centre stage. *Curr. Opin. Cell Biol.* 11, 219–225.

Marchesi VT, Furthmayr H & Tomit M (1976) The red cell membrane. *Annu. Rev. Biochem.* 45, 667–698.

Pumplin DW & Bloch RJ (1993) The membrane skeleton. *Trends Cell Biol.* 3, 113–117.

Reithmeier RAF (1993) The erythrocyte anion transporter (band 3). *Curr. Opin. Cell Biol.* 3, 515–523.

Rodgers W & Glaser M (1993) Distributions of proteins and lipids in the erythrocyte membrane. *Biochemistry* 32, 12591–12598.

Sharon N & Lis H (1995) Lectins – proteins with a sweet tooth: functions in cell recognition. *Essays Biochem.* 30, 59–75.

Sheets ED, Simson R & Jacobson K (1995) New insights into membrane dynamics from the analysis of cell surface interactions by physical methods. *Curr. Opin. Cell Biol.* 7, 707–714.

Sheetz MP (1995) Cellular plasma membrane domains. *Mol. Membr. Biol.* 12, 89–91.

Silvius JR (1992) Solubilization and functional reconsititution of biomembrane components. *Annu. Rev. Biophys. Biomol. Struct.* 21, 323–348.

Steck TL (1974) The organization of proteins in the human red blood cell membrane. *J. Cell Biol.* 62, 1–19.

Stevens TJ & Arkin IT (1999) Are membrane proteins "inside-out" proteins? *Proteins: Struct. Funct. Genet.* 36, 135–143.

Subramaniam S (1999) The structure of bacteriorhodopsin: an emerging consensus. *Curr. Opin. Struct. Biol.* 9, 462–468.

Viel A & Branton D (1996) Spectrin: on the path from structure to function. *Curr. Opin. Cell Biol.* 8, 49–55.

Vince JW & Reithmeier RA (1996) Structure of the band 3 transmembrane domain. *Cell Mol. Biol. (Noisy-le-Grand)* 42, 1041–1051.

Wallin E & von Heijne G (1998) Genome-wide analysis of integral membrane proteins from eubacterial, archaean, and eukaryotic organisms. *Protein Sci.* 7, 1029–1038.

11　小分子的膜运输和膜的电性质

膜运输的原则
载体蛋白和膜的主动运输
离子通道和膜的电特性

细胞膜的脂双层因其疏水中心而成为大多数极性分子通过的屏障。这种屏障功能至关重要，因为它允许细胞维持溶质在胞质溶胶中的浓度（与细胞外液中的浓度不同），以及细胞内各种膜包围的区域中的浓度。然而，要利用这种屏障，细胞必须进化出跨膜运输特殊水溶性分子的途径，以便摄取必需营养素、分泌代谢废物和调节胞内离子浓度。无机离子和水溶性有机小分子跨脂双层膜的运输是通过专门的跨膜蛋白来执行的，每种蛋白质负责一种特殊离子、分子或一组密切相关的离子或分子的运输。细胞也能跨膜运输大分子，甚至大颗粒，但是所涉及的机制与运输小分子所用的机制不同，这将在第 12 章和第 13 章讨论。所有生物都有大量编码运输蛋白的基因，在所有细胞中，15%～30%的膜蛋白是运输蛋白，这说明膜运输的重要性。一些特化的哺乳动物细胞将代谢总能量的 2/3 用于膜运输过程。

在本章，我们首先考虑水溶性小分子跨膜运输的一些总原则。然后，我们依次考虑介导运输的两大类膜蛋白：通过自身运动跨膜转运特异分子的载体蛋白（carrier protein）和形成一条主要允许无机小离子被动运动的狭窄亲水孔的通道蛋白（channel protein）。载体蛋白可与一个能量源偶联，用以催化主动运输。与细胞外液（表 11-1）或膜包围的细胞器内的液体相比较，胞质溶胶的成分因选择性被动渗透性和主动运输的组合而产生很大差异。细胞膜通过产生跨膜离子浓度差异以电化学梯度的形式储藏潜在的能量，用于驱动各种不同的运输过程，在电兴奋细胞中传递电信号，并产生（线粒体、叶绿体和细菌中的）大部分 ATP。我们主要集中于讨论跨质膜运输，但是跨真核细胞的其他膜也采取类似的机制，这在后续章节中将予以讨论。

表 11-1　一个典型哺乳动物细胞内外离子浓度的比较

成分	胞内浓度/(mmol/L)	胞外浓度/(mmol/L)
阳离子		
Na^+	5～15	145
K^+	140	5
Mg^{2-}	0.5	1～2
Ca^{2+}	10^{-4}	1～2
H^+	7×10^{-5}（$10^{-7.2}$mol/L 或 pH 7.2）	4×10^{-5}（$10^{-7.4}$mol/L 或 pH 7.4）
阴离子*		
Cl^-	5～15	110

* 细胞必须含有等量的正电荷和负电荷（即是电中性的）。因此，除 Cl^- 外，细胞还含有表中没有列出的许多其他阴离子；实际上，大部分细胞成分带负电（如 HCO_3^-、PO_4^{3-}、蛋白质、核酸、带磷酸基和羧基的代谢物等）。给定的 Ca^{2+} 和 Mg^{2+} 浓度是指其游离离子的浓度。细胞中的总浓度约为 20mmol/L Mg^{2+} 和 1～2mmol/L Ca^{2+}，但是这些离子大部分与蛋白质和其他物质结合，至于 Ca^{2+} 则主要储藏在各种细胞器中。

在本章的最后部分，我们主要集中于讨论神经元（神经细胞）中离子通道的功能。在这些细胞中，通道蛋白以最复杂的形式行使其功能，使神经元网络能够行使人脑所能执行的所有惊人的功能。

膜运输的原则

在这一节，我们首先描述人造无蛋白质脂双层的渗透特性。然后，我们介绍一些描述各种形式膜运输的术语，以及鉴定蛋白质及其所涉及的过程的一些策略。

无蛋白质脂双层对离子是高度不渗透的

如果有足够的时间，实际上任何一个分子都将沿着其浓度梯度跨越无蛋白质脂双层而扩散。然而，扩散效率差异很大，其原因部分是因为分子的大小，但主要还是与它在油中的相对可溶性有关。一般而言，分子越小，它在油中的可溶性就越大（即更具疏水性或非极性），也就更快地跨越脂双层膜而扩散。非极性小分子，如 O_2 和 CO_2，易溶于脂双层，因而快速跨越脂双层而扩散。不带电的极性小分子，如水或尿素，也能跨越脂双层而扩散，尽管会慢得多（图 11-1）。相反，脂双层对带电荷的分子（离子）具有高度不渗透性，无论它们有多小：这类分子的电荷和高水化度阻止它们进入脂双层的烃相。因此，人造脂双层对水的渗透性是对小离子（如 Na^+ 或 K^+）的渗透性的 10^9 倍（图 11-2）。

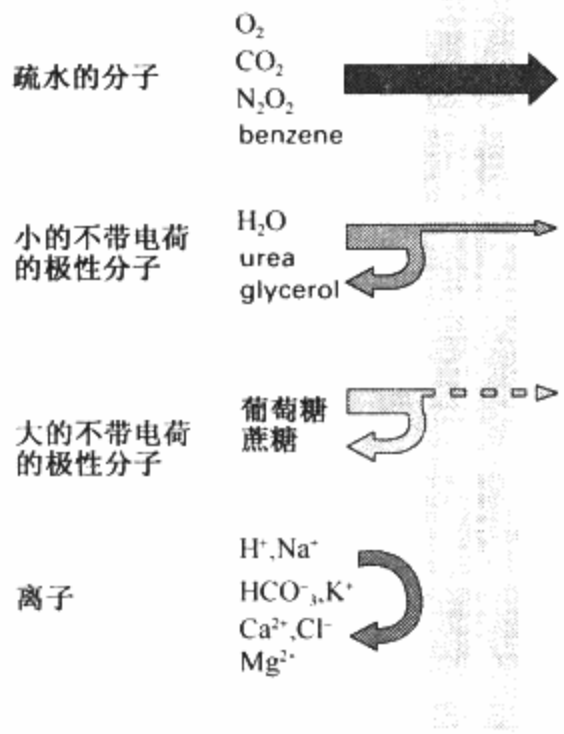


图 11-1　人造脂双层对不同类型分子的相对渗透性

分子越小越重要，与水的结合越弱，分子跨过脂双层的扩散速度就越快

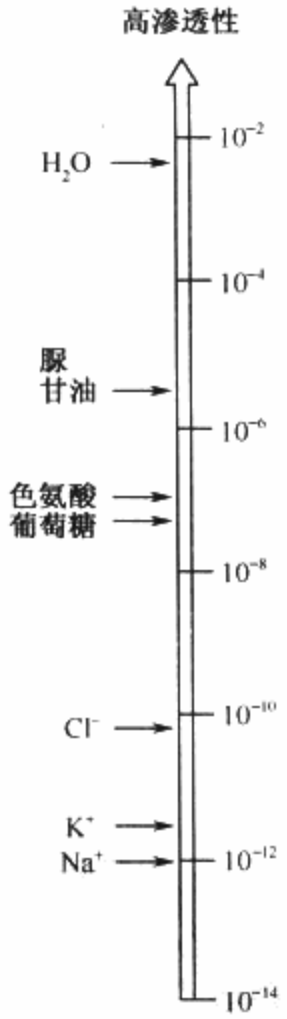


图 11-2　各种不同分子通过人造脂双层的渗透系数

溶质同脂双层的流速与膜两侧浓度的差异成正比。浓度差（以 mol/cm^3 表示）与渗透系数（以 cm/s 表示）的乘积等于溶质的流速（以 $mol/s/cm^2$ 表示）。例如，若色氨酸的浓度差为 mol/cm^3（$10^{-4}/10^{-3}$ L＝0.1mol/L），那么其流速为 10^{-4} $mol/cm^3 \times 10^{-7}$ cm/s＝10^{-11} mol/（s·cm^2），或 6×10^4 个分子/（s·μm^2）

膜运输蛋白可分为两大类：载体蛋白和通道蛋白

与人造脂双层类似，细胞膜允许水和非极性分子通过简单扩散而通过。然而，细胞膜也必须允许各种不同的极性分子通过，如离子、糖、氨基酸、核苷酸和许多细胞代谢物，只能极缓慢地跨越人造脂双层。专门的**膜转运蛋白**（membrane transport protein）负责转运这类溶质跨越细胞膜。这些蛋白质以多种形式存在于所有类型的生物膜中。每种蛋白质转运一类特殊分子（如离子、糖或氨基酸），并且常常是这类分子中的特定种类。首次提出膜转运蛋白的特异性是在 20 世纪 50 年代中期。研究发现，单基因突变使细菌转运特殊糖跨越其质膜的能力消失。现在，类似突变发现于患有影响肾、肠或其他许多细胞类型中特殊溶质转运的多种遗传病患者中。例如，患有遗传病胱氨酸尿症的个体不能从尿或肠转运特定氨基酸（包括胱氨酸——通过二硫键连接的半胱氨酸二聚体）进入血液；由此造成胱氨酸在尿中的累积导致在肾脏形成胱氨酸结石。

人们发现，所有已被详细研究的膜转运蛋白都是多跨膜蛋白，即它们的多肽链多次跨越脂双层。这些蛋白质形成连续的蛋白质跨膜通路，使特殊的亲水性溶质不直接接触脂双层的疏水内核而跨膜。

载体蛋白（carrier protein）和通道蛋白（channel protein）是膜转运蛋白的两种主要类型。**载体蛋白**也称为传递体（carrier）、通透酶（permease）或运载体（transporter），它们与待转运的特殊溶质结合，并经历一系列的构象变化，跨膜转运结合的溶质（图 11-3）。相反，**通道蛋白**与待转运溶质的相互作用弱得多。它们形成横跨脂双层、充满水的小孔。当这些小孔打开时，它们允许特殊溶质（通常是具有适当大小和电荷的无机离子）通过，因而跨越膜（图 11-3）。不足为奇的是，通过通道蛋白转运的发生频率比由载体蛋白介导的转运发生的频率高得多。

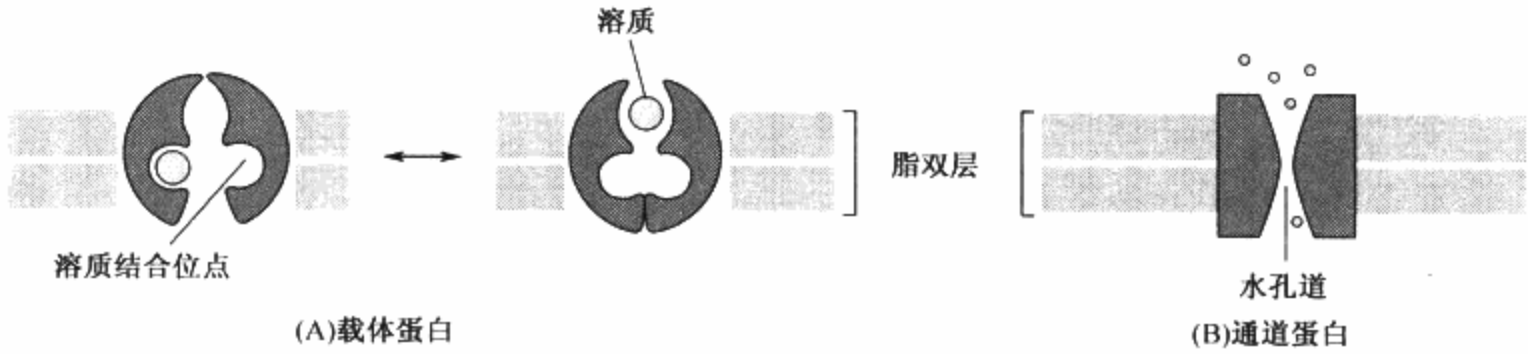


图 11-3　载体蛋白和通道蛋白

（A）载体蛋白在两种构象之间变换，结果溶质结合位点先是接近脂双层的一侧，然后接近另一侧。（B）相反，通道蛋白形成一个充满水的、跨越脂双层的孔道，特殊溶质可通过这些孔道扩散

主动运输通过与能源偶联的载体蛋白介导

所有的通道蛋白和许多载体蛋白都允许溶质被动地跨越膜，此过程被称为**被动运输**（passive transport）或**易化扩散**（facilitated diffusion）。在转运不带电荷的单个分子的情况下，仅仅是膜两侧的浓度差异，即浓度梯度（concentration gradient）驱动被动运输和确定转运方向［图 11-4（A）］。

然而，如果溶质携带净电荷，那么就是浓度梯度和跨膜电势差，即膜电位（membrane potential）影响其运输。浓度梯度和电梯度可组合起来计算每种带电溶质的净驱动力——**电化学梯度**（electrochemical gradient）［图 11-4（B）］。我们将在第 14 章更详细地讨论这一问题。实际上，几乎所有的质膜都有一个跨膜电势差（电压陡度）。相对细胞外而言，细胞内通

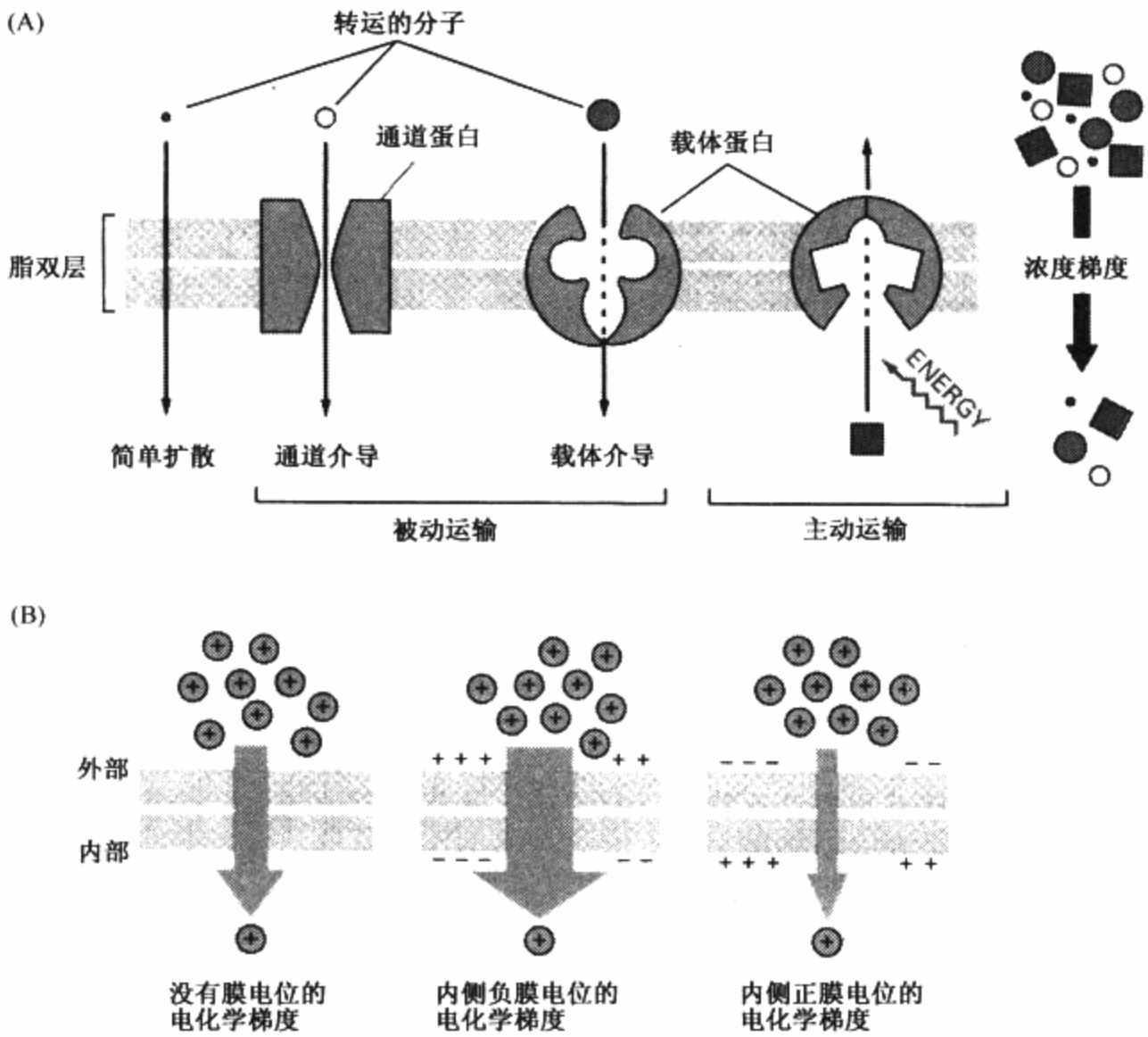


图 11-4　被动运输和主动运输的比较

（A）顺电化学梯度的被动运输时自发发生的，或者是通过简单扩散穿过脂双层，或者是通过通道载体和被动载体的易化扩散。相反，主动运输则需要输入代谢能，并且总是由聚集代谢能的载体介导，以便将溶质逆电化学梯度泵出。（B）电化学梯度由膜电位和浓度梯度综合组成，它们共同作用增加离子跨膜的驱动力（中间）或彼此对抗（右边）

常为负。这种电势差有利于正电荷的离子而不利于负电荷的离子进入细胞内。

细胞也需要转运蛋白将特定溶质逆电化学梯度主动跨膜。这一过程被称为**主动运输**（active transport），它由称之为“泵”的载体蛋白介导。在主动运输中，载体蛋白的泵活性是有方向性的，因为它与一个代谢能量源紧密偶联，如 ATP 水解或离子梯度，这一点将在后面予以讨论。因此，通过载体蛋白的转运可能是主动的或是被动的，而通过通道蛋白的运输则总是被动的。

离子载体可用作增加膜对特异离子通透性的工具

离子载体是溶解在脂双层中的疏水小分子，并增加脂双层对特殊无机离子的通透性。大多数离子载体由微生物合成（就像对付敌手或猎物的生物武器一样）。在研究人造脂双层、细胞或细胞器时，细胞生物学家将它们作为工具广泛用于增加膜的离子通透性。有两类离子载体——**可动离子载体**（mobile ion carrier）和**通道离子载体**（channel former）（图 11-5）。这两种类型的载体都是通过屏蔽被转运离子的电荷以便能贯穿脂双层的疏水中心而起作用。由于离子载体不与能量源偶联，因此它们允许离子顺着其电化学梯度而运动。

凡林霉素是可动离子载体的一个例子。它是一个环状多聚体，顺着电化学梯度转运 K^+，在膜的一侧结合 K^+，扩散通过脂双层，在膜的另一侧释放 K^+。与此类似，FCCP 是使 H^+ 选择性通过膜的一种可动离子载体，常用于消耗跨线粒体内膜的 H^+ 电化学梯度，从而阻断

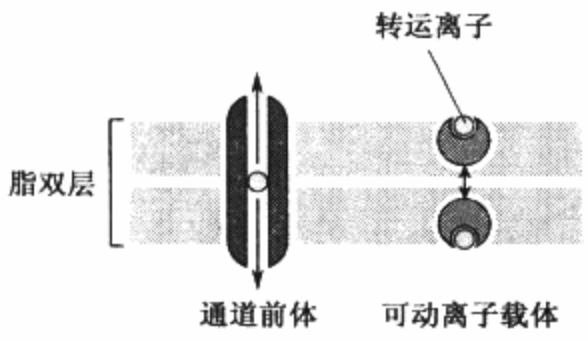


图 11-5　离子载体：通道离子载体和可动离子载体

在这两种情况下，净离子流动仅顺着电化学梯度流动

线粒体 ATP 的产生。A23187 是可动离子载体的另一个例子，仅转运二价阳离子，如 Ca^{2+} 和 Mg^{2+}。当细胞暴露于 A23187 时，Ca^{2+} 从细胞外液顺着电化学梯度进入胞质溶胶。因此，这种离子载体被广泛用于增加胞质溶胶中游离 Ca^{2+} 的浓度，从而模拟某种细胞信号转导机制（讨论见第 15 章）。

短杆菌肽 A 是通道离子载体的一个例子。这是一种由两条线性肽（每条肽链都有 15 个疏水氨基酸）组成的二聚物，肽链彼此缠绕形成双螺旋。一般认为，两个短杆菌肽二聚体横跨脂双层头尾相接聚在一起，形成一种很可能是最简单的跨膜通道，允许单价阳离子顺着电化学梯度选择性通过膜。短杆菌肽是由特定细菌产生，它杀死其他微生物是通过破坏对它们的生存至关重要的 H^+、Na^+ 和 K^+ 梯度，因而是一种有用的抗生素。

小结

脂双层对大多数极性分子都具有高度不通透性。为了转运水溶性小分子进出细胞或胞内有膜包围的区域，细胞含有各种膜转运蛋白，每种蛋白质负责跨膜转运一种或一类特殊溶质。有两类膜转运蛋白——载体蛋白和通道蛋白。二者都形成跨越脂双层的连续蛋白质通道。通过载体蛋白的转运可以是主动的、也可以是被动的，而溶质通过通道蛋白的转运却总是被动的。离子载体是微生物产生的疏水小分子，可被用作增加细胞膜对特殊无机离子渗透性的工具。

载体蛋白和膜的主动运输

载体蛋白转运溶质分子跨越脂双层的过程类似于酶-底物反应，而且载体蛋白的行为在许多方面都与酶类似。然而，与平常的酶-底物反应相反，被转运的溶质不被载体蛋白共价修饰，而是毫无改变地被转运至膜的另一侧。

每种类型的载体蛋白都有一个或多个特异的溶质（底物）结合位点。它通过经历可逆的

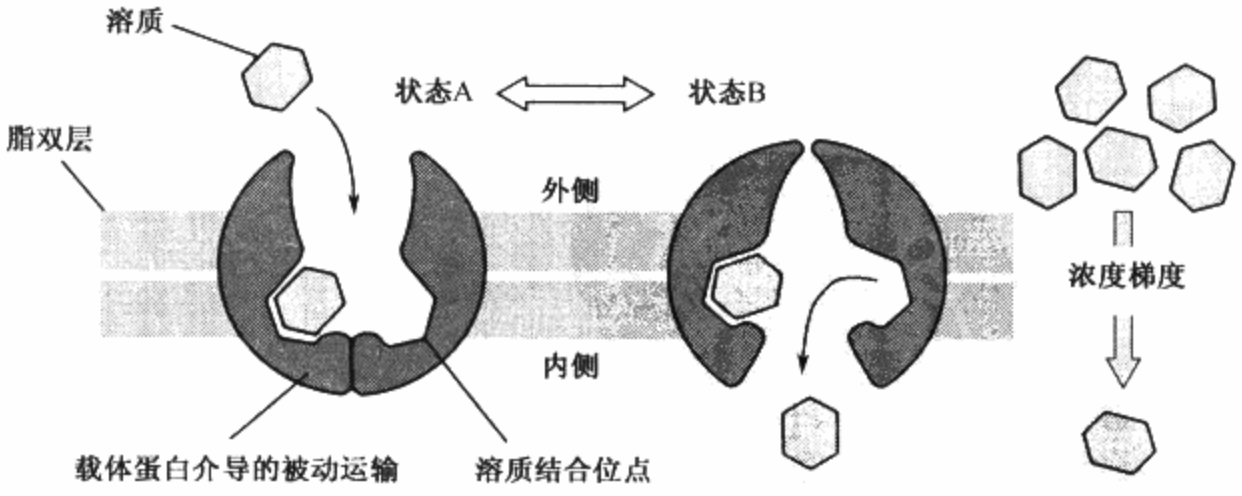


图 11-6　载体蛋白中的构象变化能介导溶质被动运输的一个模型

图中所示的载体蛋白能以两种构象状态存在：处于状态 A 时，溶质的结合位点被暴露于脂双层外侧；处于状态 B 时，相同的位点则暴露于脂双层的另一侧。在这两种状态之间的转变可随机发生。那是完全可逆的，并且不依赖于溶质结合位点是否被占据。因此，如果脂双层外侧的溶质浓度较高，那么与载体蛋白 A 构象结合的溶质会多于与 B 构象结合的溶质，结果就出现了溶质顺其浓度梯度（或者如果是离子则顺着其电化学梯度）的净运输

构象变化转运溶质跨越脂双层，这种构象变化首先将溶质结合位点暴露于膜的一侧，然后转而暴露于另一侧。载体蛋白怎样发挥作用的一个示意模型如图 11-6 所示。当载体蛋白饱和时（即当所有的溶质结合位点被占据时），转移的速率达到最大值。该速率（被称为 V_{max}）是特殊载体蛋白的特征，它反映了载体蛋白在两种构象状态之间变换的速率。另外，每个载体蛋白对其溶质都有一个特征性的结合常数 K_m，这一结合常数等于转运速率达到最大值一半时的溶质浓度（图 11-7）。就酶而言，溶质的结合可被竞争性抑制剂（竞争相同的结合位点，可能被或可能不被载体蛋白转运）或非竞争性抑制剂（与其他位点结合，特异改变载体蛋白的构象）特异阻断。

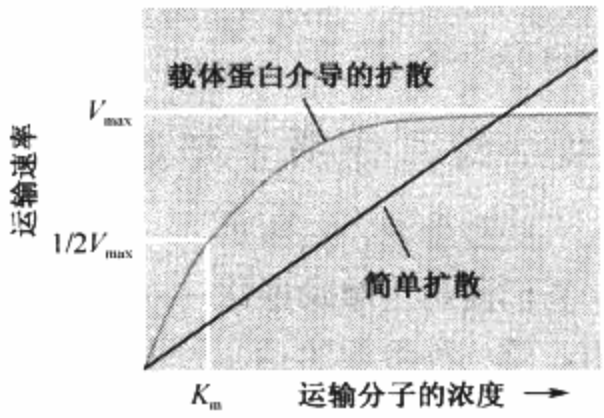


图 11-7　简单扩散和载体蛋白介导的扩散的动力学

载体蛋白的速率总是与溶质浓度成比例，然而，当载体蛋白被饱和时，其速率达到最大值（V_{max}）。当运输速度达到最大值一半时，溶质浓度近似等于载体蛋白对溶质的结合常数（K_m），这类似于酶对其底物的 K_m。该曲线图适应于运输单一溶质的载体蛋白；偶联运输两种或多种溶质的动力学（见文中）更复杂，但基本上都表现出类似的现象

我们下面将要讨论到，如图 11-6 中的模型所示，将载体蛋白与能量源偶联以便将溶质逆电化学梯度转运仅需要相对较小的修饰。细胞通过 3 种主要途径执行这种主动运输（图 11-8）：

1. 偶联载体蛋白（coupled carrier）将一种溶质的逆向跨膜转运与另一种溶质的顺向跨膜转运偶联在一起。

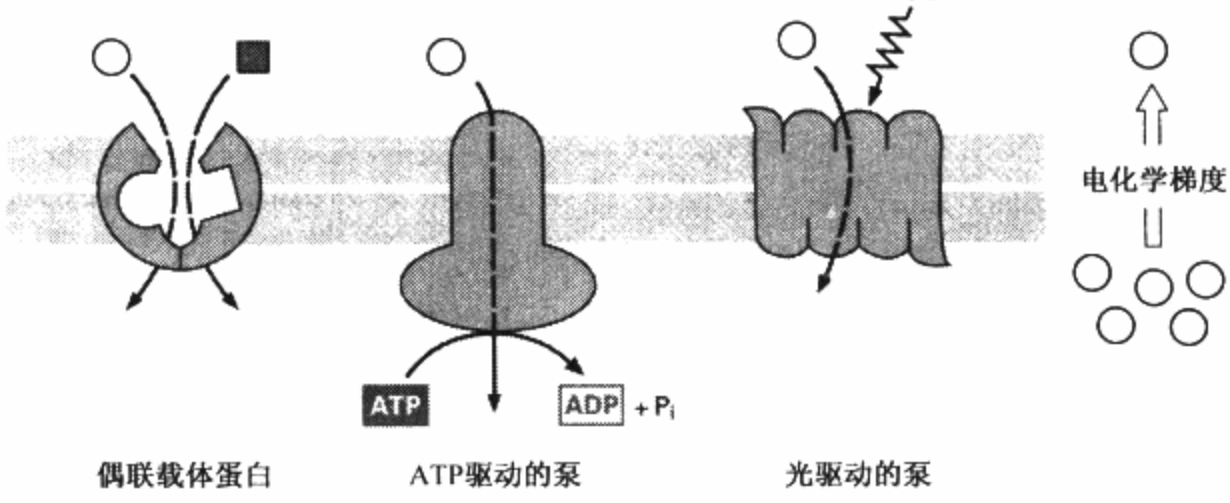


图 11-8　驱动主动运输的 3 种方式

被主动运输的分子显示为黄色，能源显示为红色

2. ATP 驱动的泵（ATP-driven pump）将逆向转运与 ATP 的水解偶联在一起。

3. 光驱动的泵（light-driven pump），主要发现于细菌细胞，将逆向转运与来自光的能量输入偶联在一起，正如细菌视紫质的情况一样（讨论见第10 章）。

氨基酸序列比较提示，在许多情况下，在介导主动运输和被动运输的载体蛋白之间的分子设计有很强的相似性。例如，一些利用储藏在跨质膜 H^+ 梯度中的能量的细菌载体蛋白在结构上类似于介导葡萄糖被动运输进入大多数动物细胞的载体蛋白。这提示各种不同载体蛋白之间的进化关系；既然小代谢产物和糖作为能源具有重要意义，那么载体蛋白超家族是一个古老的蛋白质家族就不足为奇了。

我们讨论主动运输时，首先考虑由离子梯度驱动的载体蛋白。这些蛋白质在所有细胞的小代谢产物跨膜转运中起关键作用。然后，我们讨论 ATP 驱动的泵，包括几乎在所有细胞的质膜中都存在的 Na^+ 泵。

主动运输可被离子梯度驱动

一些载体蛋白只将单一一种溶质从膜的一侧转运至另一侧，其速率由上述的 V_{max} 和 K_m 决定；它们被称为**单输送体**（uniporter）。其他载体蛋白的动力学更加复杂，起偶联载体蛋白的作用，即一种溶质的运输严格依赖于另一种溶质的运输。偶联运输或者涉及由**共输送体**（symporter）执行的同方向同时转运另一种溶质，或者涉及由**反向输送体**（antiporter）执行的逆方向同时转运另一种溶质（图 11-9）。

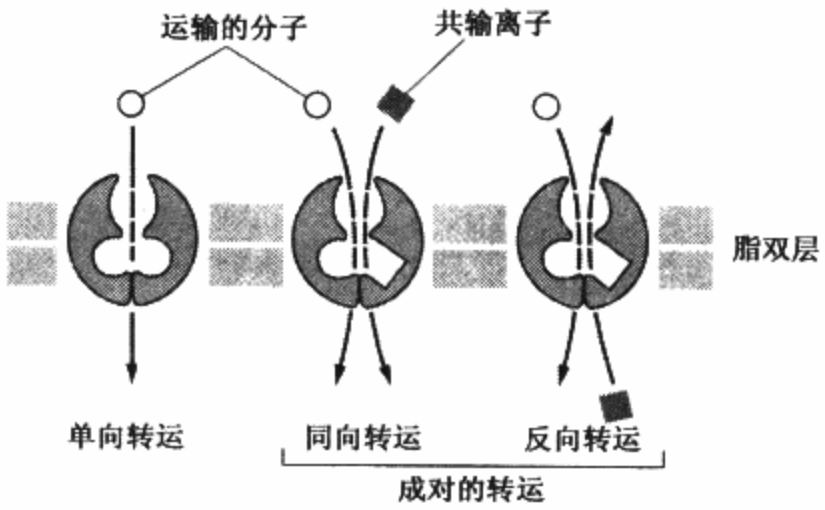


图 11-9　载体介导的 3 种运输类型

该示意图显示起单输送体、共输送体和反向输送体作用的载体蛋白

两种溶质转运的紧密偶联允许这些载体蛋白捕获储藏在一种溶质中的电化学梯度，通常是离子，用于转运另一种溶质。通过这种方式，无机离子顺着电化学梯度运动所释放的自由能被用于驱动另一种溶质逆电化学梯度运输。这种运输机制可以任一方向发挥作用；一些偶联载体蛋白以共输送体发挥作用，其他的则以反向输送体发挥作用。在动物细胞的质膜中，Na^+ 是常见的共转运离子，其电化学梯度为另一分子的主动运输提供强大的驱动力。随后，转运时进入细胞的 Na^+ 通过质膜中的 ATP 驱动的 Na^+ 泵被泵出胞外（这一点我们将在后面讨论），这是通过维持 Na^+ 梯度间接驱动转运（由于这一原因，一般认为离子驱动的载体蛋白介导原发性主动运输）。例如，肠和肾上皮细胞含有由跨膜 Na^+ 梯度驱动的多种共转运系统；每个系统将一小群相关的糖或氨基酸特异性地转运进入细胞内。在这些系统中，溶质和 Na^+ 与载体蛋白上的不同位点结合。因为 Na^+ 倾向于顺着电化学梯度转运进入细胞，所以在某种意义上，糖或氨基酸会随着 Na^+ 被“拖”入细胞内。Na^+ 的电化学梯度越大，溶质进入细胞的速率就越大；相反，如果细胞外液中的 Na^+ 浓度降低，那么溶质的转运速率就下降（图 11-10）。

就像在动物细胞的许多有膜包围的细胞器一样，在细菌和酵母中，大多数由离子梯度驱动的主动运输依赖于 H^+ 梯度而不是 Na^+ 梯度，这表明，在这些膜中，H^+ 泵占优势，而 Na^+ 泵实际上没有。例如，许多糖和氨基酸被主动运输进入细菌细胞由跨质膜的 H^+ 电化学梯度驱动。一个研究的很透彻的 H^+ 驱动的同向转运是**乳糖通透酶**（lactose permease），它转运乳糖跨越大肠杆菌的质膜。尽管通透酶的折叠结构未知，但是对突变蛋白质的生物物理研究和多方面分析提出了同向转运怎样起作用的一个详细模型。通透酶由 12 个松散折叠的跨膜 α 螺旋组

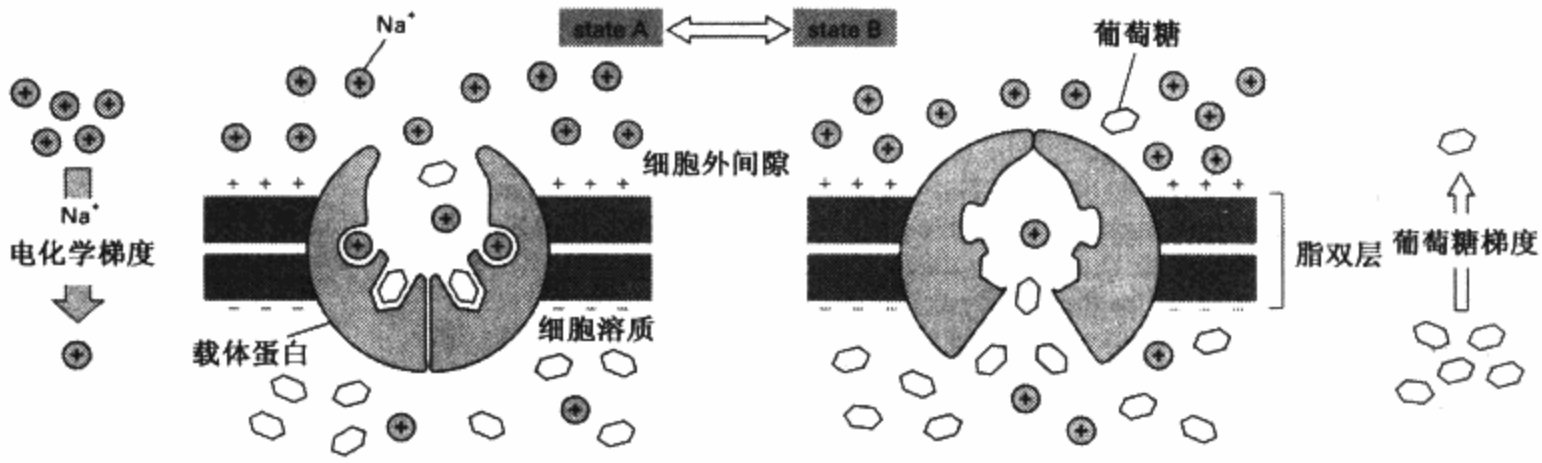


图 11-10　葡萄糖运载体被 Na^+ 梯度驱动的一种方式

如图 11-6 中所示的模型，运载体在两种状态 A 和 B 之间来回变化。在 A 状态时，蛋白质朝细胞外间隙开放；在 B 状态时，则朝胞质溶胶开放。Na^+ 和葡萄糖的结合是协同的，即任一种配基的结合都会诱导其构象变化，大大增加蛋白质对另一种配基的亲和力。由于细胞外间隙中的 Na^+ 浓度比胞质溶胶中要高得多，所以葡萄糖更可能与 A 状态的运载体结合。因此，进入细胞（通过 A→B 转变）的 Na^+ 和葡萄糖比离开细胞（通过 B→A 转变）的多得多。其总结果是 Na^+ 和葡萄糖进入细胞的净运输。注意，因为结合是协同的，如果失去两种溶质中的一种，那么另一种就不能与运载体结合。因此，如果两种溶质都与之或都不与之结合，那么运载体会在两种状态之间发生构象转变

成。发生转运循环时，一些螺旋做滑行移动，使之倾斜。这些选择性地开放和关闭螺旋之间的裂缝，结果使乳糖和 H^+ 的结合位点先是暴露于膜的一侧，然后是膜的另一侧（图 11-11）。

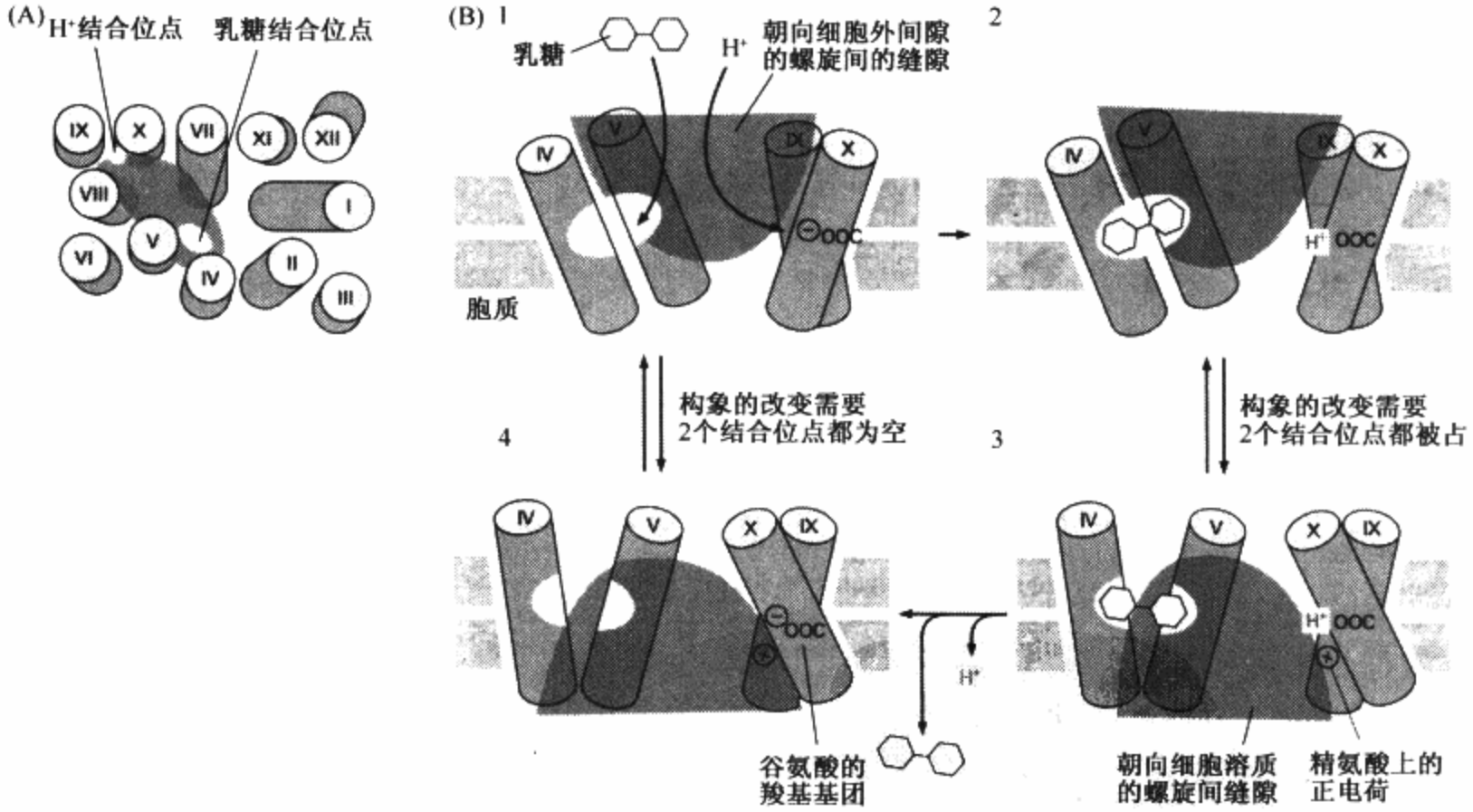


图 11-11　细菌乳糖渗透酶作用的分子机制的模型

（A）从胞质溶胶侧观察膜中 12 个预测跨膜螺旋的假定排列方式。为简洁起见，图中没有画出膜两侧连接螺旋的环。螺旋上 X 的谷氨酸与 H^+ 结合，螺旋 IV 和 V 上的氨基酸与乳糖结合。（B）在转运循环过程中，运载体在两种构象状态之间转变：在一种状态时，H^+ 和乳糖结合位点暴露于细胞外间隙（1 和 2）；在另一种状态时，暴露于胞质溶胶（3 和 4）。卸掉胞质溶胶侧的溶质（3→4）是有利的，因为乳糖结合位点受到部分破坏，并且螺旋 IX 上的精氨酸贡献的正电荷取代了来自螺旋 X 上的谷氨酸的 H^+ ［（A）改编自 H. R. Kaback and J. Wu, *Accts. Chem. Res.* 32：805～813，1999］

质膜中 Na^+ 驱动的载体蛋白调节胞质溶胶 pH

大多数大分子的结构和功能很受 pH 的影响，大多数蛋白质在特殊 pH 时发挥最佳效应。例如，溶酶体酶在溶酶体中的低 pH（约为 5）下发挥最佳效应，而胞质溶胶酶在胞质溶胶中接近中性 pH（约为 7.2）时作用最佳。因此，细胞控制其胞内区域 pH 的能力是至关重要的。

大多数细胞的质膜中有一种或多种类型的 Na^+ 驱动的反向输送体，帮助维持胞质溶胶 pH（pH_i）在 7.2 附近。这些蛋白质利用储藏在 Na^+ 梯度中的能量泵出由于渗漏或胞内的酸形成反应所产生的过多 H^+。这利用了两种机制：一种是 H^+ 直接被转运出细胞；另一种是 HCO_3^- 被带入细胞，依照反应 $HCO_3^- + H^+ \longrightarrow H_2O + CO_2$ 中和胞质溶胶中的 H^+。利用第一种机制的反向输送体之一是 Na^+-H^+ 交换体（Na^+-H^+ exchanger），它将 Na^+ 的内流与 H^+ 的外流偶联在一起。另一个利用这两种机制的反向输送体是 Na^+ 驱动的 Cl^--HCO_3^- 交换体（Cl^--HCO_3^- exchanger），它将 Na^+ 和 HCO_3^- 的内流与 Cl^- 和 H^+ 的外流偶联在一起，结果$NaHCO_3$入胞，HCl 出胞。Na^+ 驱动的 Cl^--HCO_3^- 交换体的效率是 Na^+-H^+ 交换体的 2 倍，细胞每泵入 1 分子 Na^+，就泵出 1 分子 H^+，并中和另 1 分子 H^+。如果没有 HCO_3^-（实际情况常常如此），那么这种反向输送体就是调节 pH_i 最重要的载体蛋白。这两种交换体都受到 pH_i 的调节，并随着胞质溶胶中 pH 的下降增加其活性。

不依赖于 Na^+ 的 Cl^--HCO_3^- 交换体在 pH_i 的调节中也有重要作用。与依赖于 Na^+ 的运载体类似，Cl^--HCO_3^- 交换体受到 pH_i 的调节，但是在这种情况下，HCO_3^- 常常顺其电化学梯度被转运出细胞。HCO_3^- 的外流和 Cl^- 的内流随 pH_i 的增加而增加，从而当胞质溶胶变得过碱时降低 pH_i。Cl^--HCO_3^- 交换体类似于第 10 章讨论的红细胞膜中的带 3 蛋白。在红细胞中，带 3 蛋白有利于细胞在通过非毛细血管时快速交换 CO_2。

ATP 驱动的 H^+ 泵也可用于调控许多胞内区域的 pH。正如第 13 章中所讨论的，溶酶体中的低 pH 与内体和分泌小泡中的低 pH 一样，都是通过这种 H^+ 泵来维持的，它们利用 ATP 水解的能量将 H^+ 从胞质溶胶泵入这些细胞器。

上皮细胞中载体蛋白的不对称分布是溶质跨细胞运输的基础

在上皮细胞中，如那些涉及从肠道吸收营养的细胞，载体蛋白在质膜中呈不均匀分布，从而有利于被吸收溶质的**跨细胞转运**。如图 11-12 所示，位于质膜顶端（吸收）区域与 Na^+ 相联系的共输送体将营养主动转运进入细胞，使这些溶质建立起重要的跨质膜浓度梯度。在基底外侧区域不依赖于 Na^+ 的转运蛋白允许营养顺着其浓度梯度被动离开细胞。

在这些上皮细胞中，许多细胞的质膜面积因形成数以千计的微绒毛（microvilli）而大大增加，这些微绒毛从每个细胞顶端的表面延伸成为指状的细小突起（图 11-12）。这种微绒毛能将细胞的总吸收面积增加 25 倍，从而增强其转运能力。

我们已经看到，尽管离子梯度在驱动细胞的许多重要转运过程中起关键作用，但是利用 ATP 水解能的离子泵主要负责建立和维持这些梯度，关于这一点我们将在下面予以讨论。

质膜的 Na^+-K^+ 泵是一种 ATPase

细胞内的 K^+ 浓度通常是细胞外的 10～20 倍，而 Na^+ 的浓度则与此相反（表 11-1）。这些浓度差异通过 **Na^+-K^+ 泵**或 **Na^+ 泵**来维持。这种泵存在于几乎所有动物细胞的质膜中，它们发挥反向输送体的作用，将 Na^+ 顺着电化学梯度主动排出胞外，同时将 K^+ 泵入胞内。因

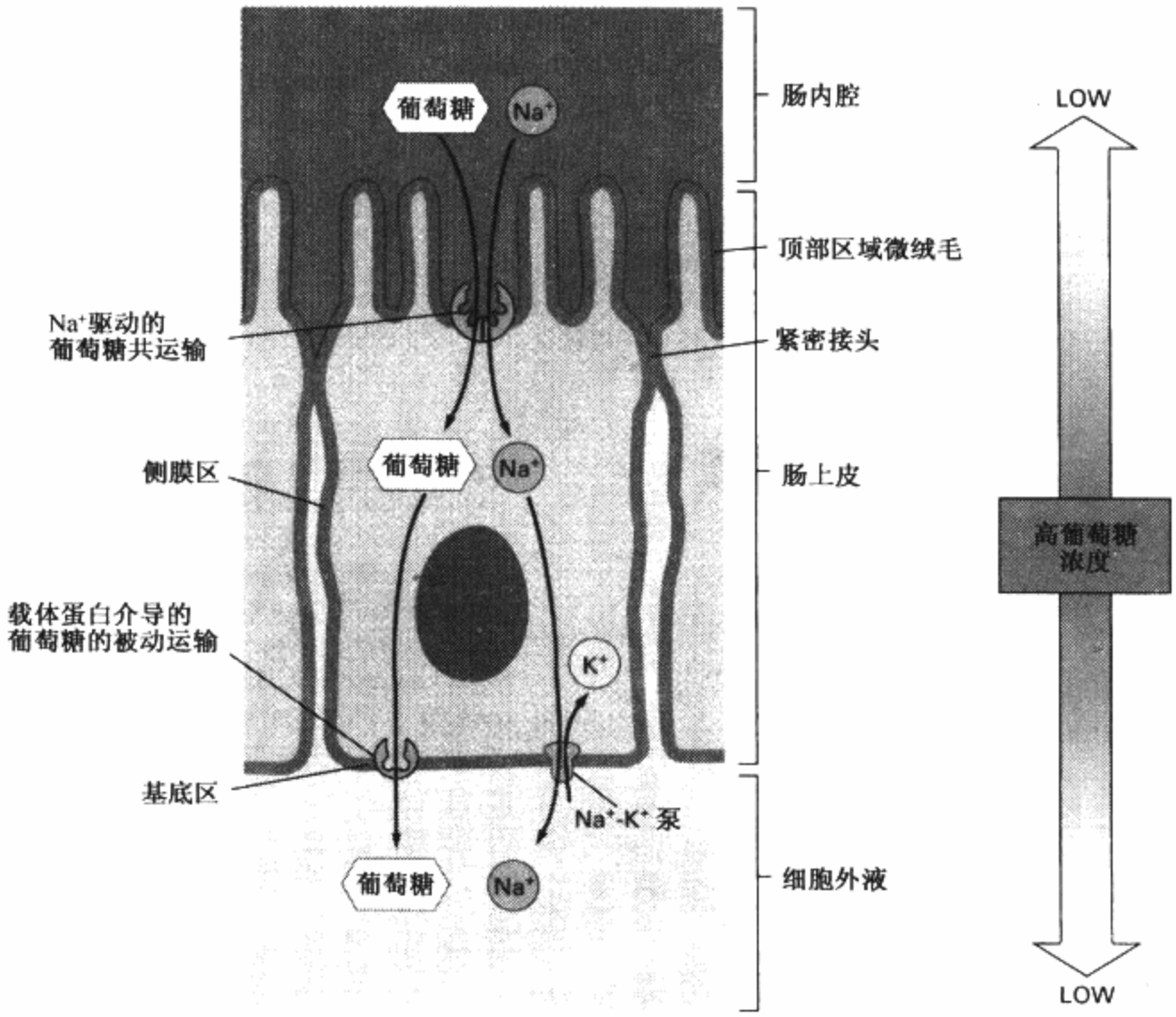


图 11-12　跨细胞转运

葡萄糖通过肠上皮细胞的跨细胞转运依赖于细胞质膜上运输蛋白的不均一分布。图中所示的过程致使葡萄糖从肠腔被转运至细胞外液（由此进入血液）。葡萄糖通过 Na^+ 驱动的葡萄糖共运输经由膜的顶端区域被泵入细胞。葡萄糖通过由基底膜和侧膜区域中的不同葡萄糖载体蛋白介导的被动运输被转运出细胞（顺着其浓度梯度）。驱动葡萄糖共运输的 Na^+ 梯度通过基底的和侧面的质膜区域中的 Na^+ 泵来维持，它们保持了胞内的低 Na^+ 浓度。邻近细胞通过不可通透的紧密连接相连，紧密连接在图中所示的过程中具有双层功能：它们阻止溶质横过细胞之间的上皮，从而允许维持跨越细胞层的葡萄糖浓度梯度，它们也充当质膜内的扩散屏障，有助于将不同的载体蛋白局限于它们各自的膜区域（参见图 10-41）

为这种泵水解 ATP 以排出 Na^+ 和泵入 K^+，所以它也被称为 **Na^+-K^+ ATPase**（图 11-13）。

我们已经看到，由泵产生的 Na^+ 梯度驱动大多数营养转运进入动物细胞，同时也在调节胞质溶胶 pH 方面起关键作用。我们在下面将会讨论到，这种泵还通过渗透效应调节细胞体积；实际上，它使许多动物细胞免于爆裂。一个典型动物细胞几乎 1/3 的能量需求都被这种泵消耗了。我们将会看到，在神经冲动传递过程中重复获得少量 Na^+ 和丢失少量 K^+ 的电活性神经细胞中，该数值达到细胞能量需求的 2/3。

Na^+-K^+ 泵的一个重要特征是其转运循环依赖于蛋白质的自磷酸化作用。ATP 的末端磷酸基团被转移给泵的天冬氨酸残基，随后被除去，解释如图 11-14 所示。因此，泵的不同状态根据磷酸基团的有无而加以区分。通过这种途径将自身磷酸化的离子泵被称为 P 型转运 ATPase（P-type transport ATPase）。它们组成一个在结构和功能上相关的蛋白质家族，包括各种 Ca^{2+} 泵和 H^+ 泵，我们将会在下面就此进行讨论。

与任何一种酶类似，Na^+-K^+ 泵可向相反方向被驱动，在这种情况下产生 ATP。当 Na^+ 和 K^+ 的梯度通过实验增加至相当高的程度，以致储藏于其电化学梯度中的能量大于 ATP 水

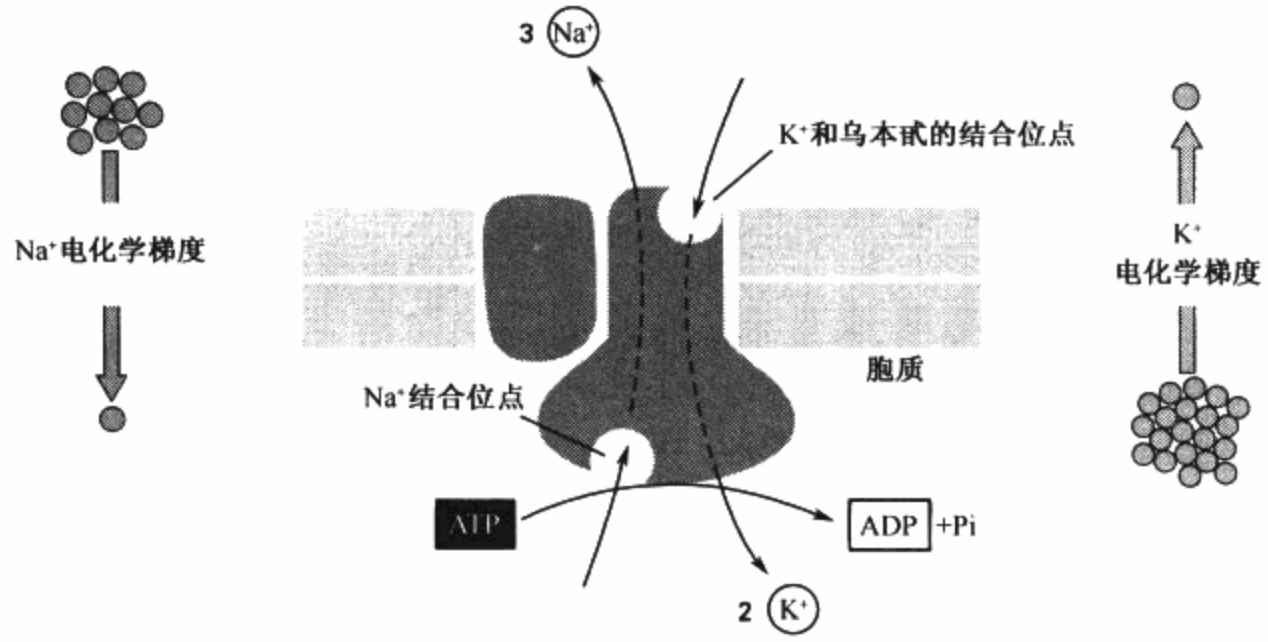


图 11-13　Na^+-K^+ 泵

该载体蛋白逆电化学梯度将 Na^+ 主动泵出胞外，并将 K^+ 主动泵入胞内。胞内每水解一分子 ATP，就有 3 个 Na^+ 被泵出胞外，2 个 K^+ 被泵入胞内。特异性抑制剂乌本甙和 K^+ 竞争泵在细胞外侧的同一位点

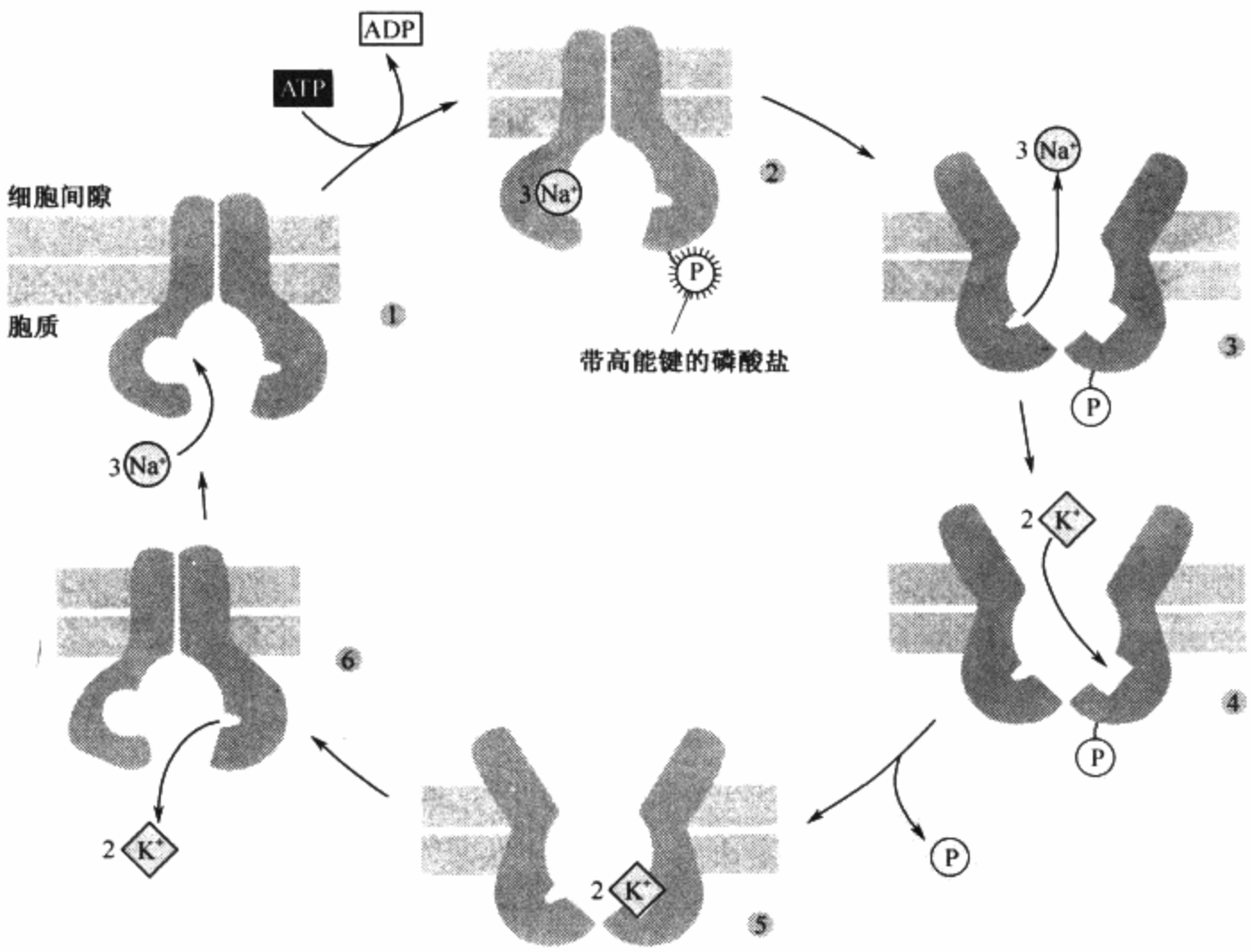


图 11-14　Na^+-K^+ 泵的泵送循环模型

①Na^+ 的结合和②随后被泵的胞质侧的 ATP 磷酸化诱发蛋白质发生构象变化，③跨膜转运 Na^+ 并将其释放至胞外。④然后，细胞外侧 K^+ 的结合和随后的去磷酸化使蛋白质回复至其最初的构象，从而跨膜转运 K^+，并将其释放至胞质溶胶。构象的这些变化类似于图 11-6 所示的 A↔B 转换，不同的是，此处 Na^+ 依赖的蛋白质磷酸化和 K^+ 依赖的蛋白质去磷酸化导致构象转变以有序的方式发生，致使蛋白质能进行有用的活动。尽管为了简单起见图中仅显示了 1 个 Na^+ 结合位点和 1 个 K^+ 结合位点，但是一般认为，一个真正的泵有 3 个 Na^+ 结合位点和 2 个 K^+ 结合位点。而且，尽管图中显示，泵仅在两种构象状态之间转换，但是有证据表明，它在泵送循环过程中经历了一系列更为复杂的构象变化

解的化学能时，这些离子顺着其电化学梯度转移，并通过 Na^+-K^+ 泵利用 ADP 和磷酸盐合成 ATP。因此，泵的磷酸化形式（图 11-14 中的步骤 2）可通过将其磷酸基团转移给 ADP（图步骤 2 到步骤 1）或者改变其构象（图步骤 2 到步骤 3）而松弛下来。自由能的全部变化是被用于合成 ATP 还是被用于将 Na^+ 排出胞外，不仅依赖于 Na^+ 和 K^+ 的电化学梯度，还依赖于 ATP、ADP 和磷酸盐的相对浓度。

一些 Ca^{2+} 泵和 H^+ 泵也是 P 型转运 ATPase

除 Na^+-K^+ 泵之外，P 型转运 ATPase 家族还包括发生信号转导事件后将 Ca^{2+} 从胞质溶胶中除去的 Ca^{2+} 泵和从胃内壁的特化上皮细胞分泌酸的 H^+-K^+ 泵。Ca^{2+} 泵尤其重要。真核细胞在胞外 Ca^{2+} 浓度极高的情况下，在其胞质溶胶中维持极低浓度的游离 Ca^{2+} 浓度（约 10^{-7}mol/L）。甚至少量的 Ca^{2+} 内流也会显著增加胞质溶胶中游离 Ca^{2+} 的浓度，Ca^{2+} 因响应胞外信号而顺着陡峭浓度梯度流动是快速传递这些跨质膜信号的一种方式（讨论见第 15 章）。因此，陡峭 Ca^{2+} 梯度的维持对细胞是重要的。Ca^{2+} 梯度的维持是通过质膜中的将 Ca^{2+} 主动转运出胞外的 Ca^{2+} 运载体。这些运载体之一是 P 型 Ca^{2+}-ATPase；另一种是反向运载体（antiporter），称为 Na^+-Ca^+ 交换体，由 Na^+ 电化学梯度驱动［图15-38（A）］。

了解得最透彻的 P 型转运 ATPase 是骨骼肌细胞肌浆网（sarcoplasmic reticulum）膜中的 **Ca^{2+} 泵**，或 **Ca^{2+}- ATPase**。肌浆网是一种特化的内质网，在肌细胞的胞质溶胶中形成管状囊网络，充当胞内 Ca^{2+} 储藏库（当一个动作电位使肌细胞膜去极化时，Ca^{2+} 通过 Ca^{2+} 释放通道从肌浆网被释放进入胞质溶胶，刺激肌肉收缩，讨论见第 16 章）。约占细胞器膜蛋白 90%的 Ca^{2+} 泵负责将 Ca^{2+} 从胞质溶胶运回肌浆网。非肌肉细胞的内质网含有类似的 Ca^{2+} 泵，但数量上更少。

已经测定了肌浆网 Ca^{2+} 泵的高分辨率三维结构。该结构和相关的真菌 H^+ 泵的分析提供了 P 型转运 ATPase 的第一种观点，即认为它们都有类似的结构。它们含有 10 个跨膜 α 螺旋，其中 3 个排列成一个跨脂双层的中央通道。在 Ca^{2+} 泵的非磷酸化状态，其中 2 个螺旋断裂形成一个空穴，该空穴可从膜的胞质溶胶侧进入，并结合 2 个 Ca^{2+}。ATP 与膜同侧的结合位点结合以及随后邻近结构域的磷酸化导致跨膜螺旋的激烈重排。这种重排破坏 Ca^{2+} 结合位点，并从膜的另一侧释放 Ca^{2+}，进入肌浆网腔（图 11-15）。

Na^+-K^+ 泵是维持渗透平衡和稳定细胞体积所必需的

由于 Na^+-K^+ 泵每泵转运 2 个正电离子入胞就会转运 3 个正电离子出胞，因此它是生电泵（electrogenic pump）。它驱动净电流跨膜，因而产生 1 个电势，细胞的内部相对于胞外呈负电性。然而，泵的这种生电效应对膜电势的贡献很少大于 10%。我们在后面将会讨论到，剩下的 90%仅仅是间接地依赖于 Na^+-K^+ 泵。

另一方面，Na^+-K^+ 泵对调节细胞容积确实有着直接而重要的影响。它调节胞内的溶质浓度，从而调节能使细胞胀大或收缩的**渗透性**（或张力）（图 11-16）。因为质膜对水的渗透性弱，所以水顺着其浓度梯度缓慢地进出细胞，此过程称为渗透作用（osmosis）。如果将细胞置于低渗溶液（hypotonic solution，即溶质的浓度低因而水的浓度高的溶液）中，那么就会出现水进入细胞的净转移，从而使得细胞肿胀而破裂（裂解）。相反，如果将细胞置于高渗溶液（hypertonic solution）中，那么它们就会因失水而收缩

(图 11-16)。许多动物细胞在其质膜中也含有特化的水通道（water channel，aquaporin），以利于渗透性的水流动。

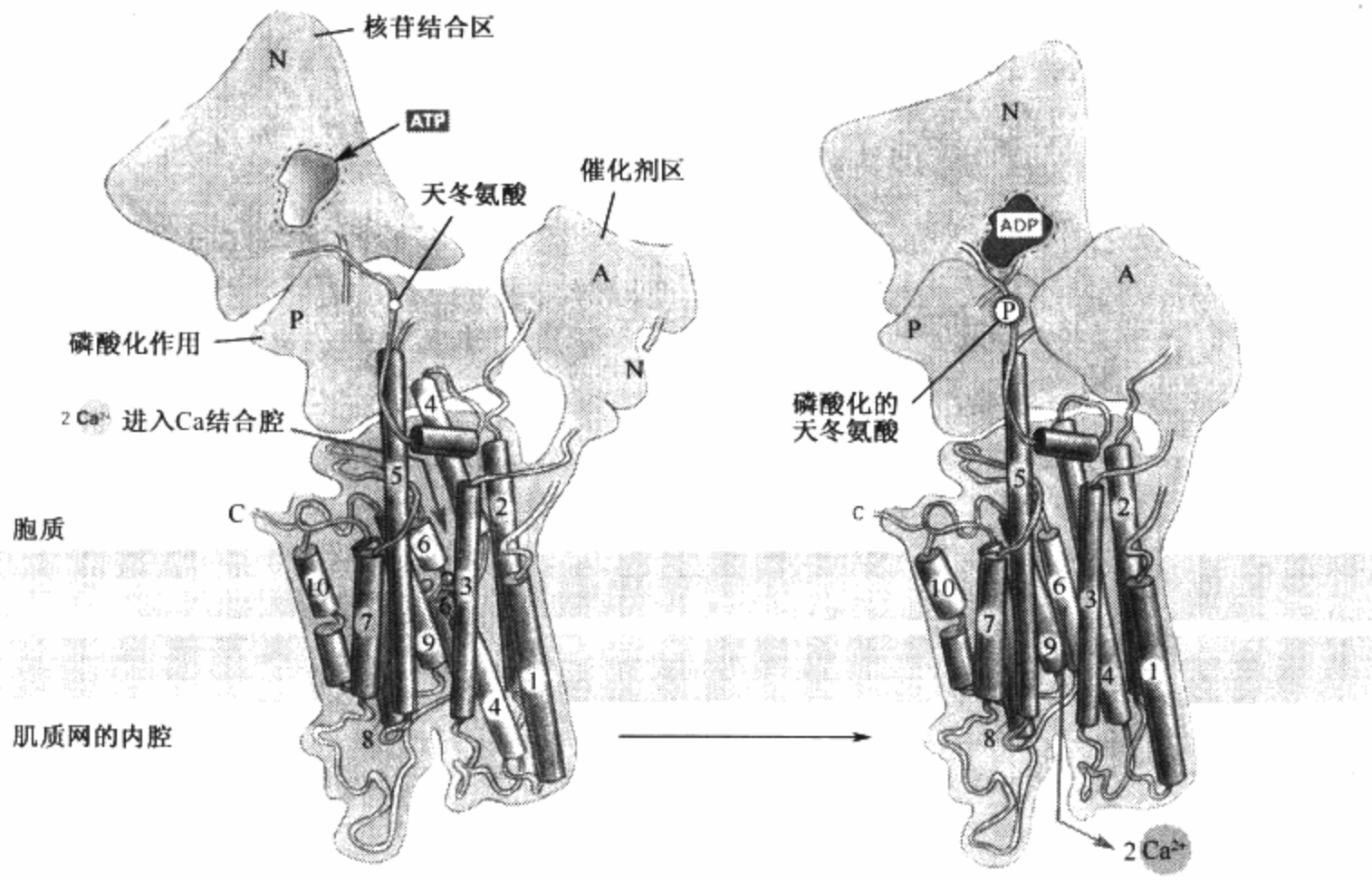


图 11-15　肌浆网 Ca^{2+} 泵怎样转移 Ca^{2+} 的一个模型

未磷酸化的 Ca^{2+} 结合状态的结构（左）是基于泵的 X 射线晶体结构。磷酸化的无 Ca^{2+} 状态的结构（右）是基于电子显微镜技术测定的低分辨率结构，跨膜螺旋的排列是推测的。在未磷酸化状态，螺旋 4 和 6 被打断，并在膜的胞质溶胶侧形成 Ca^{2+} 结合位点。ATP 的结合和水解引起显著的构象变化，从而使核苷酸结合（N）结构域和磷酸化结构域（P）紧紧靠近。一般认为，这一变化使驱动结构域（A）旋转 90°，从而导致跨膜螺旋的重排。这种重排消除了螺旋 4 和 6 中的裂隙，即废除了 Ca^{2+} 结合位点，并将膜另一侧的 Ca^{2+} 释放进入肌浆网腔（改编自 C. Toyoshima et al.，*Nature* 405：647～655，2000）

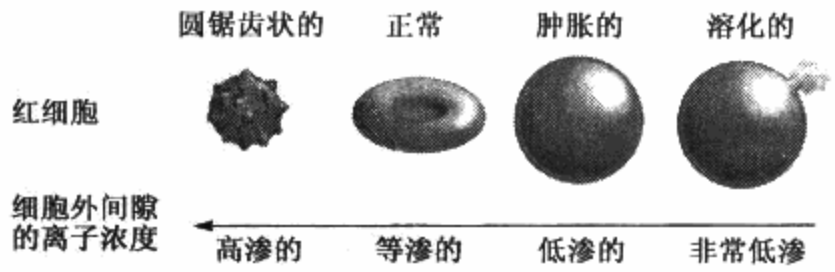
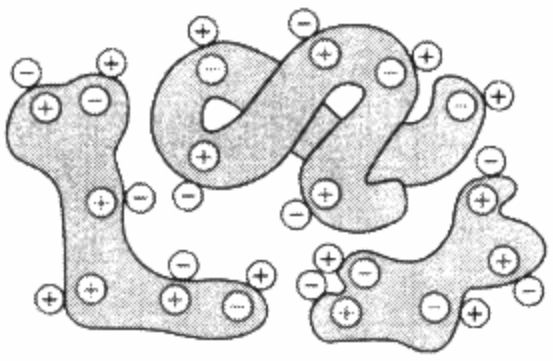


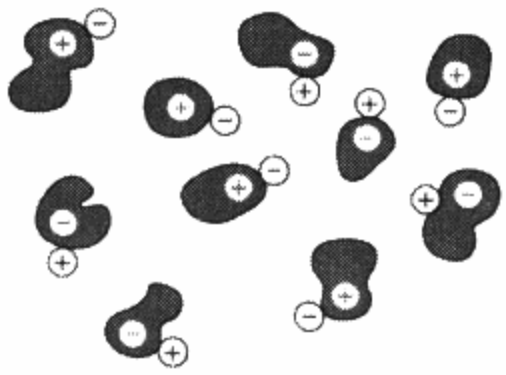
图 11-16　人红细胞对细胞外液渗透性改变的反应

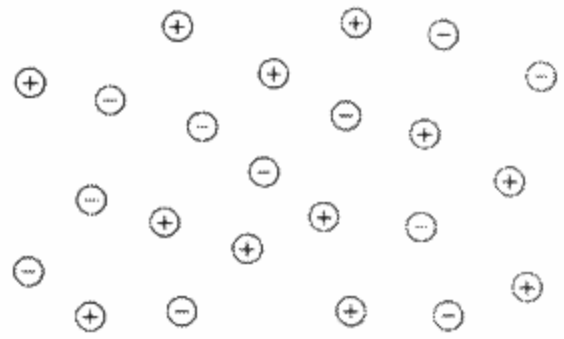
细胞随着水顺其浓度梯度的进入或者离开而始肿胀和收缩

专题 11-1　胞内水平衡：问题及其解决办法

细胞内渗透性的来源







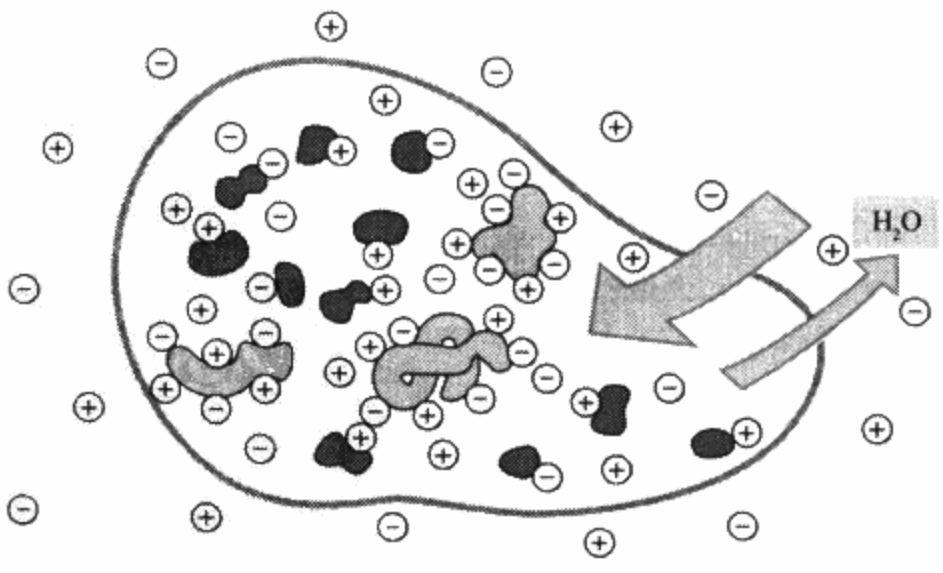
大分子本身对胞内渗透性的贡献极小，因为尽管每个分子的尺寸大，但只能以单个分子来计数，因此，与胞内小分子的数目相比，它们的分子数目相对很少。然而，大多数生物大分子带有很多电荷，它们吸引许多带相反电荷的无机离子。因其数目巨大，所以这些反离子对胞内渗透性有重要贡献。

细胞因主动运输和代谢过程而含有高浓度的无机小分子，如糖类、氨基酸和核苷酸，质膜对这些分子是不能渗透的。因为这些代谢物中大部分都带电荷，所以它们也吸引反离子。小代谢物及其反离子对胞内渗透性做出了进一步的重要贡献。

通常，细胞外液的渗透性主要是由于无机小离子。这些离子缓慢地跨膜渗漏进入细胞。如果不将它们泵出胞外，如果胞内没有与之相互作用而影响其分布的其他分子，那么它们最终将会达到平衡，即细胞内外的浓度相同。然而，胞内吸引这些离子的荷电大分子和代谢物的存在会产生唐纳效应（Donna effect）：它们使得平衡时胞内无机离子的总浓度比胞外的更大（因此对渗透性有贡献）。

问题

由于上述这些因素，一个完全不控制其渗透性的细胞，其胞内溶质浓度将会高于胞外。结果，胞外水的浓度将会高于胞内。质膜两侧水浓度的这种差异将使水通过渗透作用持续进入胞内，致使细胞裂解。

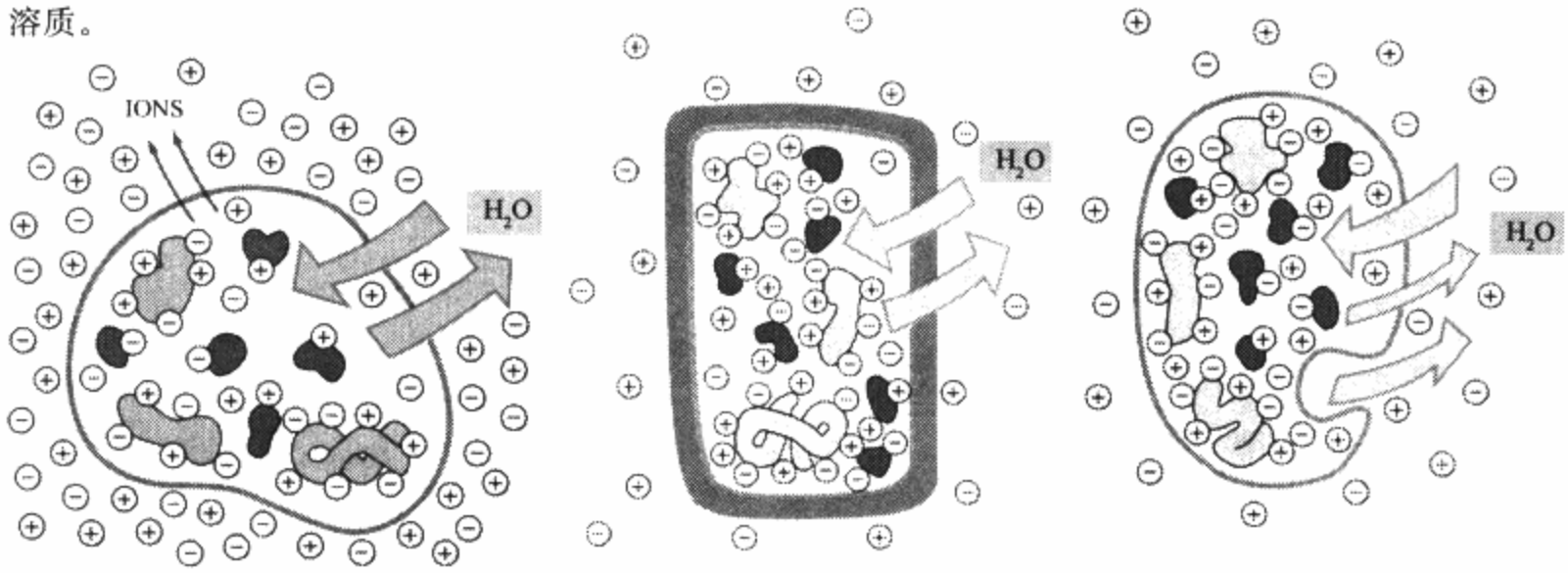


解决办法

动物细胞和细菌通过主动泵出无机离子，如 Na^+，来控制其胞内渗透性，结果胞质中的无机离子总浓度低于细胞外液，从而补偿胞内过多的有机溶质。

植物细胞通过坚硬的细胞壁阻止肿胀，因此能忍受质膜两侧的渗透性差异：内部形成的膨压在平衡时迫使尽可能多的水排出胞外。

许多原生动物通过定期将水从特殊的收缩泡中排出以避免因水而肿胀，尽管质膜两侧存在渗透性差异。



下述观察说明了 Na^+-K^+ 泵调节细胞容积的重要性：如果用抑制 Na^+-K^+ 泵的乌本甙（ouabain）处理细胞，那么许多动物细胞都会肿胀，而且常常会破裂。如专题 11-1 所解释的，细胞含有高浓度的溶质，包括大量被限制在细胞内负电荷的有机分子，即所谓的固定阴离子（fixed anion），以及电荷平衡所需要的相应阳离子。这往往导致产生很大的渗透压梯度，如果没有得到平衡，这种梯度就会将水“牵引”进入细胞。对动物细胞而言，由细胞外液中的高浓度无机离子（主要是 Na^+ 和 Cl^-）产生的一种相反的渗透压梯度抵消了这种效应。Na^+-K^+ 泵通过将渗漏进入细胞的 Na^+ 顺着其陡峭的电化学梯度排出胞外来维持渗透压平衡。Cl^- 则由于膜电势而留在胞外。

当然，细胞还可通过其他方式处理渗透压问题。植物细胞和许多细菌通过环绕质膜的半硬式细胞壁阻止其破裂。在变形虫中，渗透流入胞内的过量水被收集在收缩泡中，这些收缩泡定期将其内容物排出胞外（专题 11-1）。细菌也进化出当遭遇渗透压休克时允许快速丢失离子甚至大分子的策略。但是，对大多数动物细胞而言，Na^+-K^+ 泵是极为重要的。

合成 ATP 的膜结合酶是以相反方向起作用的转运 ATPase

细菌的质膜、线粒体内膜以及叶绿体的类囊体膜都含有转运 ATPase。不过，它们都属于**V 型ATPase**家族，在结构上与 P 型 ATPase 极为不同。F 型 ATPase 呈涡轮状结构，由多个不同的蛋白质亚基构成。我们将在第 14 章详细讨论这种类型的 ATPase。

V 型 ATPase 之所以被称为 ATP 合酶，是因为它们通常以相反方向起作用，即不是通过 ATP 水解来驱动离子运输，而是通过跨膜 H^+ 梯度驱动自 ADP 和磷酸盐合成 ATP。H^+ 梯度的产生或者是在氧化磷酸化（在需氧细菌和线粒体中）的电子传递步骤中，或者是在光合作用过程中，或者是通过盐杆菌（*Halobacterium*）的光激活 H^+ 泵（细菌视紫质）。

与 P 型转运 ATPase 类似，ATP 的合成能以任何一个方向起作用：当跨膜电化学梯度降至低于某一域值时，它们将水解 ATP，并跨膜转运 H^+。结构上与 F 型 ATPase 相关的是一个不同的 V 型 ATPase 家族。某些细胞器，如溶酶体、突触泡和植物液泡，含有 V 型

ATPase，将 H^+ 泵入细胞器，由此造成内环境的酸化。因此，V 型 ATPase 通常是转运蛋白质，而不是合成 ATP。

ABC 运载体是膜转运蛋白的最大家族

我们讨论的最后一类载体蛋白是一个具有重要临床意义的转运 ATPase 家族，纵使它们在真核细胞中的正常功能仅仅是初见端倪。这些有待鉴定的蛋白质首先在细菌中发现。我们已经提到，所有细菌的质膜都含有利用跨膜 H^+ 梯度将多种营养泵入细胞的载体蛋白。许多细菌还有转运 ATPase，利用 ATP 水解释放的能量摄入某些小分子。在有双层膜的细菌（图 11-17），如大肠杆菌中，转运 ATPase 定位于内膜，而且还存在一种捕获营养物并将它们传输给运载体的辅助机制（图 11-18）。

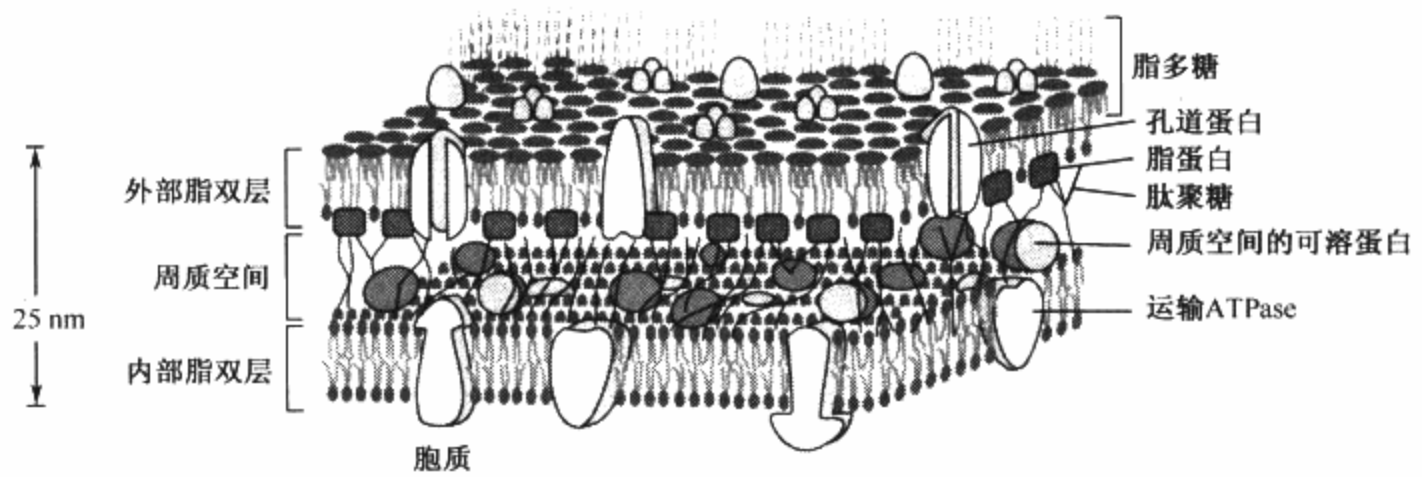


图 11-17　大肠杆菌双层膜的一个小区域

内膜是细胞的质膜。内外脂双层膜之间是由蛋白质和多糖组成的、高度多孔而坚硬的肽聚糖，它构成细菌的细胞壁。它附着在外膜的脂蛋白分子上，充满周质间隙（图中仅显示少量肽聚糖）。这一间隙也含有许多种可溶性蛋白质分子。顶部的虚线（显示为绿色）代表特殊脂多糖分子的多糖链，形成外膜的胞外单层；为清楚起见，途中仅显示少数几条链。具有双层膜的细菌被称为革兰氏阴性菌，因为它们不滞留在革兰氏染色法中用到的深蓝色染料。具有单层膜的细菌（但有更厚的细胞壁），如葡萄球菌和链球菌，滞留蓝色染料，因此被称为革兰氏阳性菌；它们的单层膜类似于革兰氏阴性菌的内膜（质膜）

细菌质膜中的转运 ATPase 属于已知最大且种类最多的转运蛋白家族。这一家族被称为**ABC 运载体超家族**（ABC transporter superfamily），因为该家族的每个成员都含有 2 个高度保守的 ATP 结合盒（图 11-19）。ATP 的结合导致 2 个 ATP 结合结构域二聚化，ATP 的水解导致二聚体的解离。一般认为，胞质溶胶结构域的这些结构变化被传递给跨膜束，从而驱动构象变化循环，交替暴露底物结合位点于膜的一侧或另一侧。ABC 运载体通过这种方式利用 ATP 的结合和水解跨脂双层转运分子。

在大肠杆菌中，ABC 运载体由 78 个基因编码（占细菌基因的 5%），动物细胞则包含更多的基因。一般认为每种运载体特异于一个或一类特殊的底物，但是由该超家族转运的底物种类繁多，包括氨基酸、糖、无机离子、多糖、肽，甚至蛋白质。细菌的 ABC 运载体被同时用于输入和输出，而在真核细胞中那些曾被描述过的运载体似乎大多数转用于输出。ABC 运载体也催化脂质从脂双层的一侧翻转至另一侧，因而在膜的生物发生和维持中具有重要作用，这将在第 12 章中予以讨论。当底物是脂质或总体上是疏水分子时，用于结合底物的结合位点必须暴露于与脂双层的疏水内部相接触的运载体的表面。

实际上，第一个被鉴定的真核 ABC 运载体是由于它们能将疏水药物从胞质溶胶中泵出而发现的。其中之一是**多向性抗药蛋白**（multidrug resistance protein，MDR protein），它在人癌细胞中超量表达能使细胞同时对多种化学上无关的广泛用于癌症化疗的细胞毒类药物产生抗

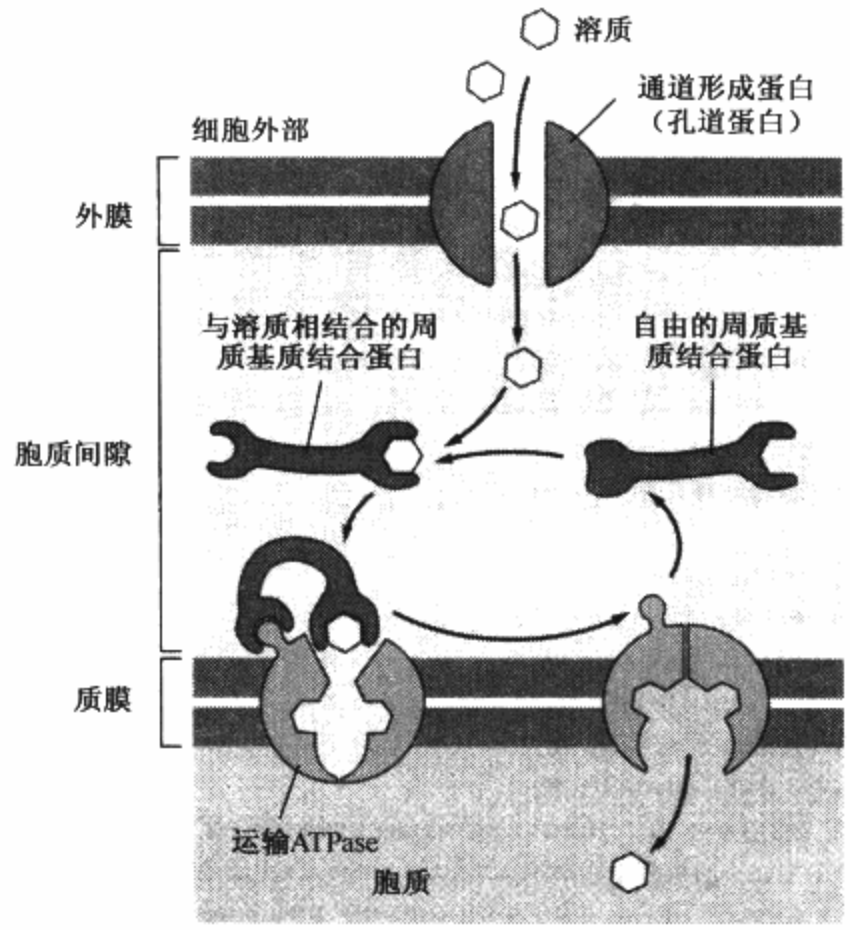


图 11-18　具有双层膜的细菌中与转运 ATPase 有关的辅助转运系统

溶质通过外膜中的通道形成蛋白（孔道蛋白）扩散，并与周质基质结合蛋白结合。结果，基质结合蛋白经历构象变化，使之能与质膜中的 ABC 运载体结合，然后，ABC 运载体接过溶质，并在一个由 ATP 水解驱动的反应中跨越脂双层主动转运溶质。为简单起见，图中省去了肽聚糖；它的多孔结构允许基质结合蛋白和水溶性分子通过简单扩散而通过

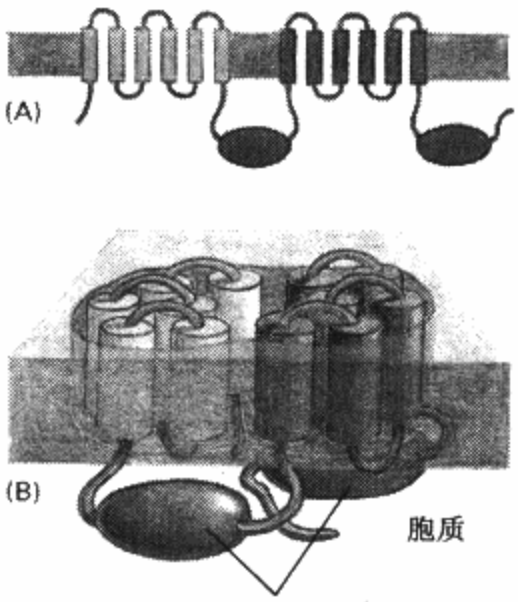


图 11-19　一个典型的 ABC 运载体

（A）拓扑学线图。（B）膜中多肽链的假定排列方式。运载体由 4 个结构域组成：两个高度疏水的结构域，每个结构域都有 6 个推测的形成易位通路的跨膜片段，两个 ATP 结合催化结构域（或盒）。在某些情况下，运载体的两部分由单一多肽（如图所示）形成，而在另一些情况下，它们由 2 个或多个独立的多肽形成，组装成一个相似结构（图 10-32）

药性。用其中任何一种药物进行治疗都能导致筛选出超量表达 MDR 转运蛋白的细胞。运载体将药物泵出细胞，从而降低其毒性，并对各种各样的治疗因素产生抗性。一些研究显示，多达 40%的人类癌症发展出多向性抗药，从而使之成为与癌症做斗争中出现的一个主要障碍。

在引起疟疾的原生动物恶性疟原虫（*Plasmodium falciparum*）中存在一种有同等危害的相关现象。有 2 亿多人感染这种寄生虫，它仍然是人类死亡的一个主要病因，每年导致一百多万人死亡。疟疾的控制因产生抗疟药氯喹（chloroquine）的抗性而受到阻碍，并且研究显示，抗药性恶性疟原虫扩增了编码将氯喹泵出胞外的 ABC 运载体的一个基因。

在酵母中，ABC 运载体跨质膜输出一种交配信息素（一种由 12 个氨基酸组成的多肽）。在大多数脊椎动物细胞中，在内质网（ER）膜中的 ABC 运载体蛋白质降解产生的各种各样的多肽从胞质溶胶主动运输至 ER。这是在免疫系统（讨论见第 24 章）的细胞监视中具有重要意义的一条途径中的第一步。转运的蛋白质片段进入 ER 后，最终被转运至细胞表面，并且它们在此被细胞毒 T 淋巴细胞识别。如果这些片段貌似异己的话，细胞毒 T 淋巴细胞就会杀死该细胞

（正如如果这些片段源自潜伏在胞质溶胶中的病毒或其他微生物所发生的那样）。

然而，ABC 家族的另一个成员是通过研究常见遗传性疾病囊性纤维化（cystic fibrosis）而被发现的。这种疾病是由编码 ABC 运载体的一个基因中的一个突变而引起的，这种 ABC 运载体的功能是调节上皮细胞质膜中的 Cl^- 通道。在 27 个白种人中有一人携带编码该蛋白质的一个突变基因；在 2500 人中有一人该基因的 2 个拷贝发生突变，因而引起这种疾病。ABC 运载体怎样调节 Cl^- 的跨膜传递仍不明确。

小结

载体蛋白与特殊溶质结合，并通过其构象变化跨脂双层转运这些溶质，这些构象变化使溶质结合位点暴露于膜的一侧，继而是另一侧。一些载体蛋白只顺浓度梯度转运一种溶质，而另一些载体蛋白则能以泵的形式逆其电化学梯度转运一种溶质，它们利用 ATP 水解、另一种溶质（如 Na^+ 或 H^+）的顺浓度梯度转运或光提供的能量以有序的方式驱动转运所必需的一系列构象变化。载体蛋白归属于少数几个家族。每个家族都包含一些具有类似氨基酸序列的蛋白质，一般认为，这些氨基酸序列是从一个共同的祖先蛋白进化而来并以类似的机制起作用。P 型转运 ATPase 家族，包括无处不在的 Na^+-K^+ 泵，就是一个重要的例子；这些 ATPase 在转运循环中相继使其自身磷酸化和去磷酸化。ABC 运载体超家族是膜转运蛋白最大的家族，在临床上尤其重要。它不仅包括产生囊性纤维化的蛋白质，还包括在癌细胞和引起疟疾的寄生虫中产生抗药性的蛋白质。

离子通道和膜的电特性

与载体蛋白不同，通道蛋白形成亲水性跨膜孔道。有一类在所有动物中发现的通道蛋白在 2 个邻近细胞之间形成间隙连接（gap junction）；每个质膜对形成通道的贡献相等，这种通道连接 2 个细胞的胞质。这些通道将在第 19 章讨论，在此不做进一步讨论。间隙连接以及细菌、线粒体和叶绿体外膜的孔道形成蛋白——孔道蛋白（porin）（讨论见第 10 章）都有相对大且允许渗透的孔道，如果这些孔道直接将细胞的内部与细胞外间隙相连接，那将是灾难性的。实际上，许多细菌毒素确实会那样做，以便杀死其他细胞（讨论见第 25 章）。

相反，动植物细胞中连接胞质溶胶和细胞外部的大多数通道蛋白必然会有能开合的高度选择性窄孔道。因为这些蛋白质尤其与无机离子的运输有关，所以它们被称作**离子通道**（ion channel）。就转运效率而言，通道优于载体蛋白，因为每个开放的孔道每秒钟能使多达 100 万个离子通过——其速率比任何一个已知载体蛋白介导的最大运输速率高 10^5 倍。然而，通道不能与能量源偶联而进行主动运输，因此它们介导的运输总是被动的（顺梯度）。因而，离子通道的功能是允许特殊无机离子（主要是 Na^+、K^+、Ca^{2+} 或 Cl^-）顺着其电化学梯度快速扩散跨越脂双层。我们将会看到，调控离子通过这些通道而内流的能力对许多细胞功能的发挥是必不可少的。特别地，神经细胞（神经元）专门利用离子通道，我们将讨论它们怎样利用多种通道接收、传导和传递信号。

离子通道是离子选择性的并在开关两种状态之间转换

有两个重要特性区分离子通道和简单含水孔道。首先，它们表现出离子选择性（ion selectivity），允许一些无机离子通过，而另一些离子则不能通过。这一特性暗示其孔道必须足

够窄，从而迫使正在渗入的离子与通道壁紧密接触，以便仅有适当大小和电荷的离子能通过。正在深入的离子不得不释放它们的大部分或所有结合水分子，以单列通过通道的最窄部分，该处被称为选择性滤过装置（selectivity filter）；这限制了它们的通过速率。因此，随着离子浓度的增加，离子通过通道的流量按比例增加直至达到最大值（饱和）。

离子通道和简单含水通道的第 2 个区别是，离子通道并不持续开放。相反，它们是门控通道，允许短暂开放，然后又关闭（图 11-20）。在大多数情况下，门因相应一种特殊的刺激而开放。已知引起离子通道开放的主要刺激类型是跨膜电压的改变（电压门控通道，voltage-gated channel）、机械应激（机械门控通道，mechanically gated channel）或配基的结合（配体门控通道，ligand-gated channel）。配体可以是细胞外介质，尤其是神经递质（递质门控通道，transmitter-gated channel），或是细胞内介质，如离子［离子门控通道（ion-gated channel）］或核苷酸［核苷酸门控通道（nucleotide-gated channel）］（图 11-21）。另外，许多离子通道的活性受到蛋白质磷酸化和去磷酸化的调节；这种类型的通道调节将在第 15 章与核苷酸门控通道一起讨论。而且，随着（化学的或电的）刺激的持续，大多数通道进入一种“脱敏的”或“失活的”关闭状态，在这种状态下，它们难以进一步开放，直至除去刺激，关于这一点我们将在后面予以讨论。

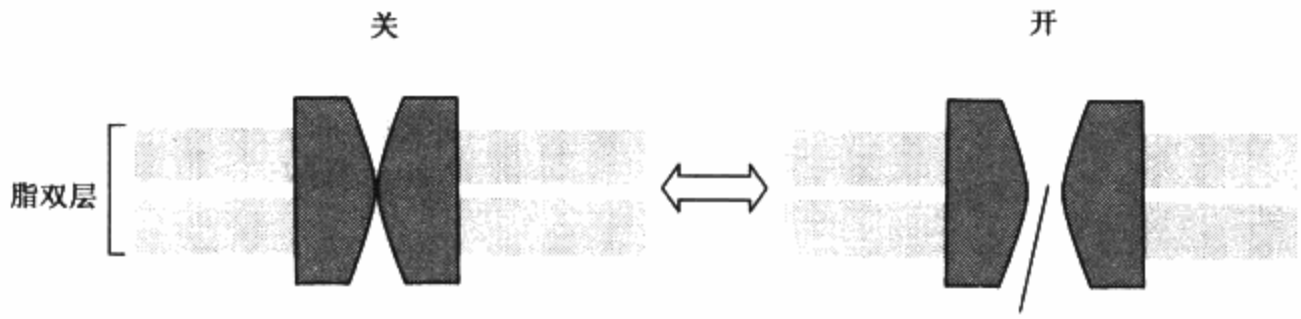


图 11-20　在闭合和开放构象之间变动的一个典型离子通道

此处以横截面显示的通道蛋白仅在“开放”构象状态形成一个跨脂双层的亲水孔道。一般认为，极性基团沿着孔道壁排列成行，而疏水氨基酸侧链则与脂双层（图中为显示）相互作用。孔道在某一区域（选择性滤过装置）变窄至原子尺度，通道的离子选择性主要由此处决定

迄今已有超过 100 种离子通道被描述，而且新的离子通道正在增加。它们负责肌肉细胞的电兴奋性，在神经系统中，它们介导几乎全部形式的电信号。通常，单个神经元可能含有 10 种或更多离子通道，定位于质膜的不同区域。但是，离子通道并不局限于电兴奋细胞。它们存在于所有动物细胞中，也在植物细胞和微生物中发现：如它们在含羞草中传播合叶反应，允许单细胞草履虫（*Paramecium*）在受到攻击后改变方向。

也许，最常见的是那些主要用于渗透 K^+ 的离子通道。这些通道发现于几乎所有动物细胞的质膜中。有一组重要的 K^+ 通道甚至在未受刺激或“静息”细胞中也开放，因而有时称之为 **K^+ 漏电通道**（K^+ leak channel）。尽管该术语适用于多种不同的 K^+ 通道（依赖于细胞类型），它们都有一个共同的目的。它们使质膜对 K^+ 的渗透能力比对其他离子的渗透能力大得多，因而在维持所有质膜的膜电势中具有至关重要的作用。

动物细胞中的膜电势主要依赖于 K^+ 漏电通道和跨质膜 K^+ 梯度

当膜一侧的阳离子略微超过阴离子而在另一侧略显不足，致使膜两侧的电荷存在差异时，就产生了**膜电势**。这种电荷差异可能产生于活性的生电泵和离子的被动扩散。如我们将在第

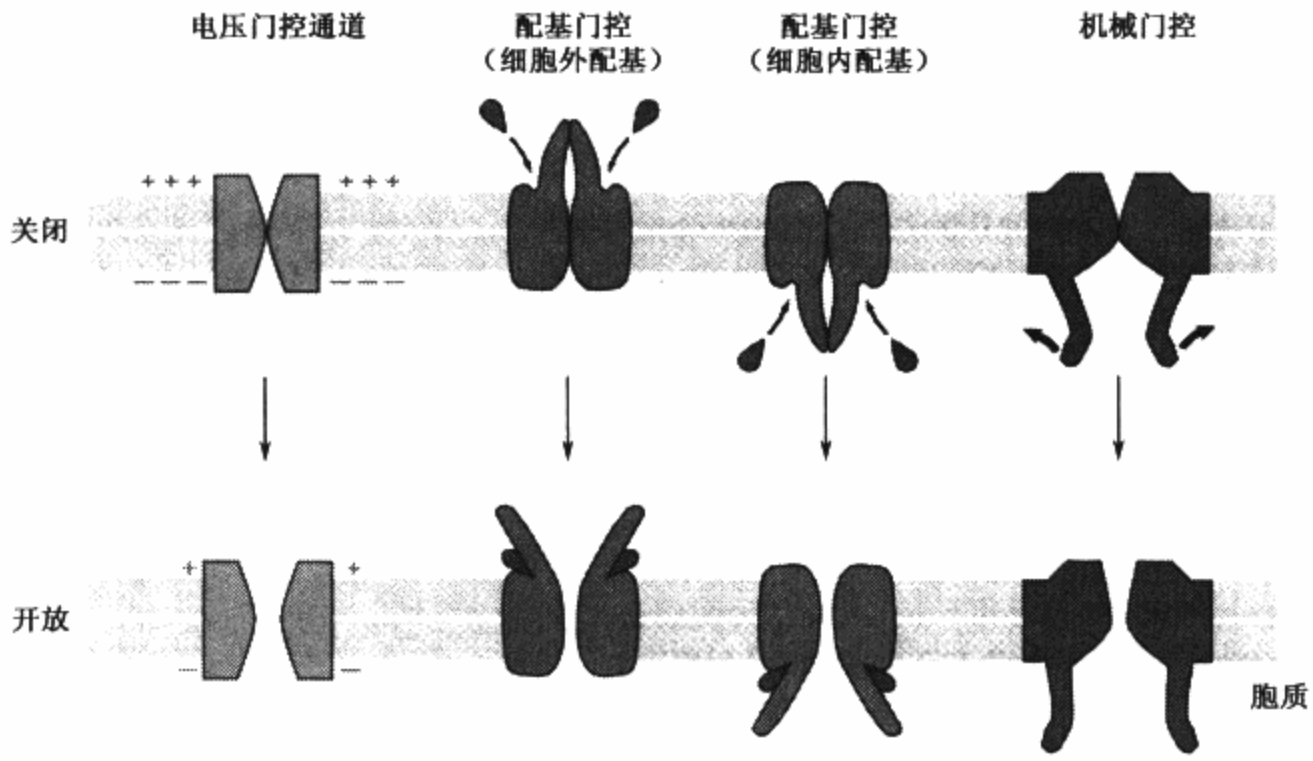


图 11-21　离子通道的门控

此图显示了开放离子通道的数种不同刺激。机械门控通道常常有胞质伸展装置，将通道和胞质骨架连在一起（图中未显示）

14 章所讨论的，线粒体的大部分膜电势产生于线粒体内膜的 H^+ 生电泵。在植物和真菌中，生电泵同样产生大部分跨质膜电势。然而，在典型的动物细胞中，被动离子运动对跨质膜电势的贡献最大。

前面解释到，Na^+-K^+ 泵通过保持胞内 Na^+ 的低浓度来帮助维持动物细胞膜两侧的渗透平衡。因为细胞内的 Na^+ 很少，所以其他阳离子必定大量存在于此，以平衡细胞的固定阴离子（限制于细胞内带负电的有机分子）所携带的电荷。这种平衡作用主要由 K^+ 来执行，K^+ 由 Na^+-K^+ 泵主动泵入细胞，它也可通过质膜中的 K^+ 漏电通道自由进出。因为这些通道的存在，K^+ 差不多达到平衡，由吸引 K^+ 进入细胞的过量负电荷所施加的电动力平衡了 K^+ 顺着其浓度梯度渗漏的倾向。膜电势是这种电动力的表现形式，它的平衡值可根据 K^+ 浓度梯度的陡度进行计算。以下论证可能有助于阐明这一点。

假设开始时没有跨质膜电压梯度（膜电势为零），但是细胞内的 K^+ 浓度高而细胞外的 K^+ 浓度低。K^+ 在其浓度梯度的驱动下，将倾向于通过 K^+ 漏电通道离开细胞。随着 K^+ 的迁出产生了一种未被平衡的负电荷，从而产生一个电场或膜电势，这将有助于对抗 K^+ 的进一步外流。当膜电势达到某个值，对 K^+ 的电驱动力正好平衡了它的浓度梯度的影响时，即当 K^+ 的电化学梯度为零时，K^+ 的净外流停止。尽管 Cl^- 也在膜两侧保持平衡，但是膜电势使这些离子大部分留在胞外，因其电荷为负。

这一平衡条件（没有离子跨质膜的净流动）定义了这种理想化细胞的**静息膜电位**（resting membrane potential）。一个简单而非常重要的公式［**能斯特方程**（Nernst equation）］定量表达这种平衡条件，就像在专题 11-2 中所解释的那样，如果已知内外离子浓度的比值，那么利用该方程可能计算出理论静息膜电位。然而，由于一个静息细胞的质膜并不只是能使 K^+ 和 Cl^- 透过，所以真正的静息膜电位通常并不是正好等于能斯特方程根据 K^+ 或 Cl^- 的预测值。

专题 11-2　能斯特方程的推导

能斯特方程和离子流

任何离子通过膜通道蛋白的流动都是由该离子的电化学梯度来驱动的。这一梯度代表两种影响的综合：电压梯度和该离子的跨膜浓度梯度。当这两种影响彼此正好平衡时，该离子的电化学梯度为零，没有离子通过通道的净流动。达到平衡时的电压梯度（膜电位）被称为该离子的平衡电位。它可通过下面推导的方程——能斯特方程来计算。

能斯特方程为

$$V=\frac{RT}{zF}\ln\frac{C_o}{C_i}$$

式中，V 表示平衡电位（胞内电位减去胞外电位），以伏特表示；C_o 和 C_i 表示分别为离子的胞外浓度和胞内浓度；R 表示气体常数 [2 cal/（mol·K)]；T 表示绝对温度（K)；F 表示法拉第常数 [2.3×10^4 cal/(V·mol)]；z 表示离子的价（电荷)；ln 表示底 e 的对数。

能斯特方程推导如下：

溶液中的分子（溶质）完全是由于分子的随机运动而倾向于从高浓度区域向低浓度区域移动，并最终达到平衡。因此，顺浓度梯度的移动伴随有利的自由能变化（$\Delta G<0$)，而逆浓度梯度的运动则伴随不利的自由能变化（$\Delta G>0$)（自由能在第 784 页的专题 14-1 引入并加以讨论。）每摩尔溶质跨质膜移动的自由能变化（$\Delta G_{conc}<0$）等于 $-RT\ln C_o/C_i$。如果溶质是离子，跨膜进入一个细胞，胞内相对于胞外的电势差为 V，那么将会引起额外的自由能变化（每摩尔移动的溶质）$\Delta G_{volt}=zFV$。当浓度梯度和电势差梯度正好达到平衡时，$\Delta G_{volt}+\Delta G_{conc}=0$，膜两侧的离子分布达到平衡。于是，

$$zFV-RT\ln C_o/C_i=0$$

因此，

$$V=\frac{RT}{zF}\ln\frac{C_o}{C_i}=2.3\frac{RT}{zF}\log_{10}\frac{C_o}{C_i}$$

对于单价离子，

$2.3\frac{RT}{zF}$=58mV（20℃时）和 61.5mV（37℃时)。

于是，对于 37℃ 的单价离子，$C_o/C_i=10$ 时，$V=+61.5$mV，而 $C_o/C_i=1$ 时，$V=0$。

例如，K^+ 的平衡电位（V_K）为 $61.5\log_{10}([K^+]_o/[K^+]_i)$ mV（一个典型细胞为 −89 mV，其中 $[K^+]_o=5$mV，$[K^+]_I=140$mV)。在 V_K 时，没有 K^+ 的跨膜净流动。相似的，当膜电位达到 $61.5\log_{10}([Na^+]_o/[Na^+]_i)$，即 Na^+ 的平衡电位（V_{Na}）时，没有 Na^+ 的净流动。

对于任何特殊的膜电位 V_M，将特殊类型能够的离子驱出细胞的净力与 V_M 和离子的平衡电位的差成比例：因此，对 K^+，它是 V_M-V_K；对 Na^+，它是 V_M-V_{Na}。

与胞内的离子总数相比，在膜附近形成电荷层的离子数目是很少的。例如，6000 个 Na^+ 跨过 $1\mu m^2$ 的膜，就有足够的电荷将膜电位改变约 100mV。因为在一个典型细胞（胞质体积为 $1\mu m^3$）内约有 3×10^7 个 Na^+，所以这种电荷运动对跨膜离子浓度梯度的影响通常是微不足道的。

当 Na^+-K^+ 泵停止时，静息电位仅缓慢下降

由于建立膜电势而必须跨质膜迁移的离子数目很少。因此，可以认为膜电势起因于电荷的运动，这种电荷运动实际上并不影响离子浓度，致使膜两侧正负离子数目的偏差极小（图 11-22)。而且，电荷的这些运动通常极其迅速，仅耗费几毫秒或更短。

在一个真实的细胞中，如果 Na^+-K^+ 泵突然失活，那么考虑膜电势将会出现何种情况是具有启发意义的。立即就会出现膜电势的轻微下降。这是因为这种泵是一种生电泵，当处于

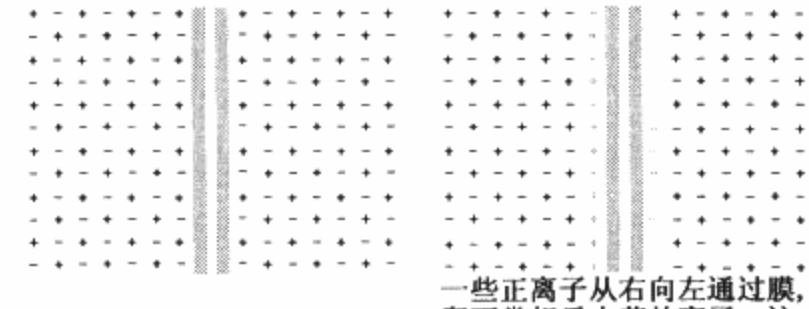


图 11-22　膜电位的离子基础

小股离子流所携带的电荷足以引起膜电位的巨大变化。产生膜电位的离子位于紧靠膜表面的一薄层（<1 nm），通过对膜另一侧带相反电荷的反离子的电荷吸引力而得以维持。对于一个典型的细胞而言，每平方厘米膜上 1 微库仑电荷从膜的一侧转移至另一侧大约使膜电位改变 1V。这意味着，例如，在一个直径为 10μm 的球形细胞中，将膜电位改变 100mV 必须流出胞外的 K^+ 数目约仅为胞质溶胶中 K^+ 总数的 1/100 000

活性状态时，它通过泵入 2 个 K^+ 的同时泵出 3 个 Na^+ 而对膜电势做出直接的微小贡献。然而，该泵的关闭不会消除静息电位的主要组成部分，因为静息电位主要是由上述的 K^+ 平衡机制产生的。只要胞内保持低 Na^+ 浓度和高 K^+ 浓度——通常为许多分钟，膜电势的这一组成部分就会持续存在。但是质膜对所有小离子，包括 Na^+，多少都会有渗透性。因此，如果没有 Na^+-K^+ 泵，那么通过泵建立起来的离子梯度最终将被消耗殆尽，通过 K^+ 漏电通道的扩散而建立起来的膜电势也将下降。随着 Na^+ 的进入，渗透平衡被打破，因而水渗入细胞（第 699～700 页的专题 11-1）。如果细胞不破裂，那么它最终会达到一个新的静息状态，此时膜两侧的 Na^+、K^+ 和 Cl^- 都将处于平衡状态。这种状态下的膜电势比正常细胞中 Na^+-K^+ 泵处于活动状态时的膜电势要小得多。

动物细胞在静息状态时质膜两侧的电势差根据生物体和细胞类型的不同在 −20～−200mV变化。尽管 K^+ 梯度总是对这一电势产生主要影响，但是其他离子的梯度（和离子泵的失衡效应）也有重要影响：膜对给定离子的渗透性越大，用于驱动该离子达到平衡值的膜电势就更大。结果，膜对离子的渗透性的变化可使膜电势发生显著变化。这是细胞的电兴奋性与离子通道的活动相关联的主要原理之一。

为了弄清楚离子通道怎样选择其离子以及它们如何开关，我们必须知道它们的原子结构。第一个被结晶用于 X 射线衍射研究的离子通道是细菌的 K^+ 通道。它的结构细节彻底改变了我们对离子通道的理解。

细菌 K^+ 通道的三维结构证明离子通道如何起作用

离子通道将巧妙的离子选择性和高效传递结合在一起的惊人能力长期以来使科学家迷惑不解。例如，K^+ 漏电通道转递 K^+ 的效率是 Na^+ 的 10 000 倍，而这两种离子是毫无特色且直径相似（分别为 0.133nm 和 0.095nm）的球体。K^+ 通道的孔内单个氨基酸替换可致使它丧失离子选择性，并造成细胞死亡。正常的选择性不能用孔道大小来解释，因为 Na^+ 比 K^+ 小。而且，高效传递与通道具有选择性的高亲和力 K^+ 结合位点不相符，因为 K^+ 与这类位点的结合会大大降低其传递效率。

当一个细菌 K^+ 通道通过 X 射线衍射晶体分析法而被测定时，这一谜团才得以解开。该通道由 4 个完全相同的跨膜亚单位组成，这些亚单位共同形成一个跨膜的中央孔道（图 11-23）。带负电的氨基酸集中在孔道的胞质溶胶侧入口处，一般认为，它们吸引阳离子，排

斥阴离子，从而使得通道具有阳离子选择性。每个亚单位形成 2 个跨膜螺旋，并向膜外倾斜，共同形成一个圆锥体，朝向胞外的一端宽阔，K^+ 从此处离开通道。连接 2 个跨膜螺旋的多肽链形成一个短螺旋，即孔螺旋（pore helix），和一个关键环，伸入圆锥体的宽阔部分，形成**选择性滤过装置**。4 个亚单位的选择性环形成一个短小的刚性狭窄孔道，其多肽骨架上的羰基氧原子沿着孔道排列。因为所有已知 K^+ 通道的选择性环都具有类似的氨基酸序列，所以它们很可能形成一个非常相似的结构。晶体结构显示，在选择性滤过装置中有 2 个成单行排列的 K^+，相隔约 8Å。一般认为，2 个离子之间的相互排斥帮助它们穿越孔道进入细胞外液。

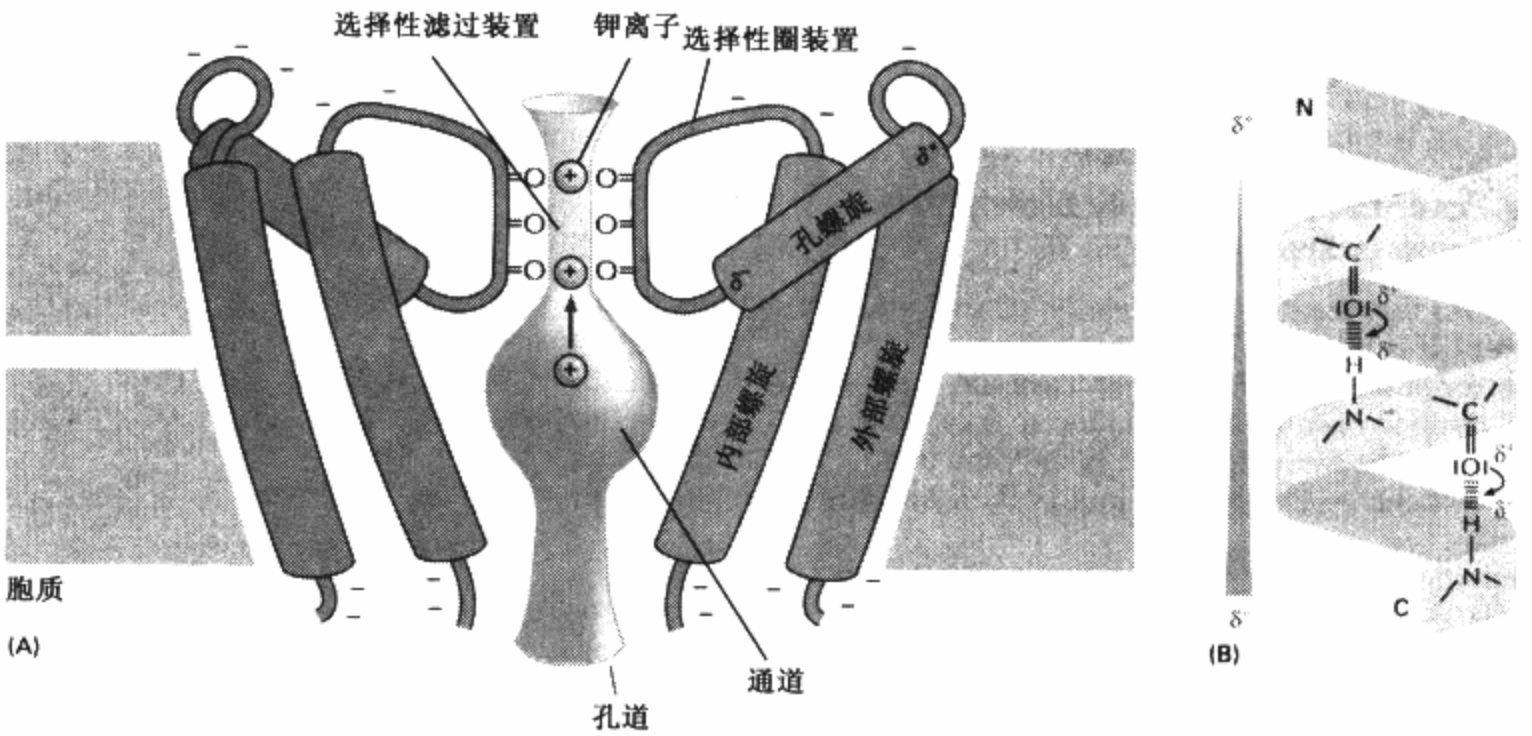


图 11-23　细菌 K^+ 通道的结构

（A）图中仅显示了 4 个完全相同的亚单位中的 2 个。从胞质溶胶侧看，孔道打开，延伸进入膜中部的孔腔中。孔腔通过允许 K^+ 保持水化状态而便于运输，即使它们仅是部分跨膜。狭窄的选择性滤过装置使孔腔连通胞外。羰基氧沿着选择性滤过装置的壁成行排列，形成脱水 K^+ 的瞬时结合位点。孔道中 K^+ 的位置是通过将通道蛋白的结晶浸泡在含铷离子的溶液中而确定的，铷离子有更高的电子密度，但仅略微大于 K^+；根据通道中 K^+ 和铷离子的衍射图的差异，可将离子的位置计算出来。2 个 K^+ 位于选择性滤过装置中，第三个离子则位于孔腔的中央，并通过与孔道螺旋有更多负电荷的末端发生电相互作用而被稳定。4 个孔道螺旋的末端精确地指向孔腔的中央。从而引导 K^+ 进入选择性滤过装置。（B）由于连接一个 α 螺旋的邻近转角的氢键（红色）具有极性，所以每个 α 螺旋都有一个沿其轴线的电偶极子，并携带一个有较多负电荷的 C 端（δ^-）和一个有较多正电荷的 N 端（δ^+）[（A）改编自 D. A. Doyle et al., *Science* 280: 69～77, 1998]

选择性滤过装置的结构解释了通道巧妙的离子选择性。如果一个 K^+ 进入滤过装置，它必须丢失几乎所有的结合水分子，而与排列于选择性滤过装置内的羰基氧相互作用，这些羰基氧以精确的距离严格间隔开，用以容纳一个 K^+。相反，Na^+ 不能进入滤过装置，因为羰基氧距较小的 Na^+ 太远，因而不能补偿与（进入时所必需的）水分子丢失有关的能量消耗（图 11-24）。

细菌 K^+ 通道的结构研究指出这些通道可能的开关方式。形成选择性滤过装置的环是刚性的，在通道开关时不会改变其构象。相反，当通道关闭时，沿孔道其余部分排列的内外跨膜螺旋发生重排，使得孔道在胞质溶胶端像光阑一样收缩（图 11-25）。尽管孔道不会完全关闭，但是存留下来的小开口被疏水氨基酸侧链塞满，从而阻断离子的进入。

充分利用离子通道的细胞是神经元。在讨论它们怎样利用之前，我们必须换个话题，简要回顾一下一个典型的神经元是被怎样组织起来的。

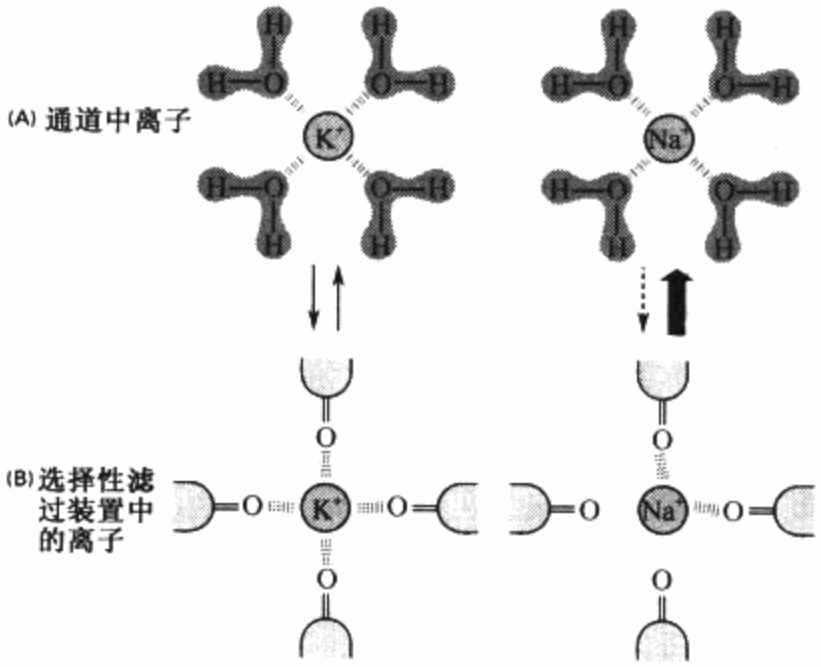


图 11-24　K^+ 通道中的选择性滤过装置的 K^+ 选择性

该图显示位于孔道的孔腔（A）和选择性滤过装置（B）中的 K^+ 和 Na^+（横切面观）。在孔腔中，离子与水结合。在选择性滤过装置中，羰基氧所处位置正好容纳一个脱水的 K^+。K^+ 的脱水需要能量，这些能量是通过离子与羰基氧（充当水分子的代用品）的相互作用而复得的能量被精确平衡的。因为 Na^+ 太小而不能与氧相互作用，所以它只能以巨大的能量消耗为代价而进入选择性滤过装置。因此，该装置对 K^+ 具有高度特异的选择性（改编自 D. A. Doyle et al.，*Science* 280：69～77，1998）

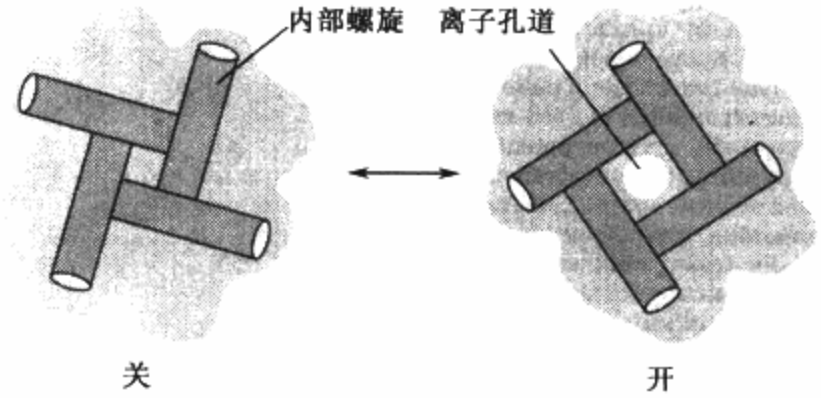


图 11-25　门控细菌 K^+ 通道的一个模型

如图是通道的横切面观。为了采用闭合构象，沿选择性滤过装置的胞质溶胶侧的孔道（图 11-22）排列的 4 个内部跨膜螺旋发生重排，从而关闭胞质溶胶侧通往通道的入口（改编自 E. Perozo et al.，*Science* 285：73～78，1999）

神经细胞的功能依赖于它的纵长结构

神经元或**神经细胞**（nerve cell）的基本功能是接收、执行和传递信号。为了执行这些功能，神经元常常延伸得很长。例如，人的单个神经细胞从脊髓延伸至足部肌肉可长达 1m。每个神经元都由一个含有许多细小突起的细胞体（含细胞核）组成，这些突起呈放射状向外延伸。通常，一根长的**轴突**远离细胞体朝远处的靶执行信号，数根较短的分支状**树突**像天线一样自细胞体延伸，从而提供更大的表面面积，接收来自其他神经细胞的轴突发出的信号（图 11-26）。细胞体本身也接收信号。一个典型的轴突在其远端分成许多分支，同时将信号传递给许多靶细胞。同样，树突的分支程度也非常大——在某些情况下，单个神经元足以接收多达 100 000 个输入信号。

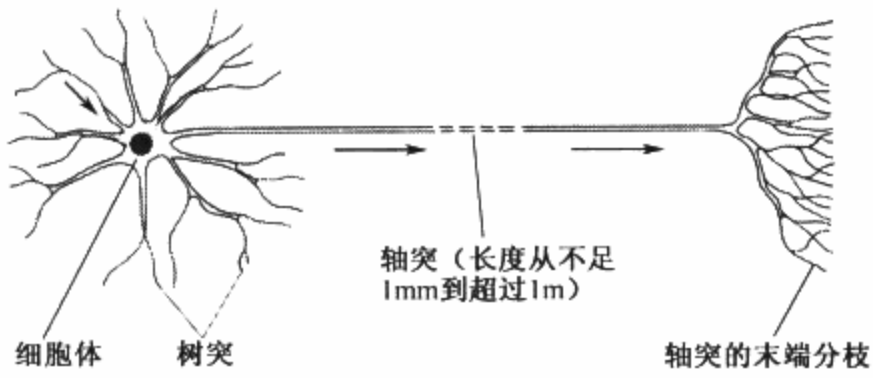


图 11-26　一个典型的脊椎动物神经元

箭头标明信号传递的方向。单个神经元引导信号远离细胞体，而多个树突则接收来自另一个神经元轴突的信号。神经末梢终止于另一个神经元的树突或细胞体，或者另一细胞类型，如肌细胞或腺细胞

尽管不同类型神经元所携带的信号有各种不同的含义，但其信号的形成始终是相同的，包含神经元质膜两侧的跨膜电势的改变。细胞的一部分所产生的电干扰传递至细胞的其他部分，因而产生细胞通讯。这种干扰随着距干扰源的距离的增加而减弱，除非能量在传递的同时被放大。这种减弱在短距离范围内是不重要的；实际上，许多小神经元被动的传递信号，而不会将信号放大。然而，对于长距离通讯而言，被动传递信号是不够的。因此，较大的神经元采用主动的信号传递机制，这是它们最显著的特征之一。超过某一阈强度的电刺激会引发生物电活性沿着神经元的质膜迅速传播，同时，生物电活性自始至终通过自动放大得以维持。被称为**动作电位**（action potential）或神经冲动（nerve impulse）的电兴奋的行波能将一种信息从神经元的一端传播至另一端，速度高达每秒 100m 或更快，而信号却并不衰减。我们在下面将会看到，动作电位是电压门控阳离子通道的特性直接造成的。

电压门控阳离子通道在电兴奋细胞中产生动作电位

所有电兴奋细胞（不仅神经元，而且还有肌肉细胞、内分泌细胞和卵细胞）的质膜都含有**电压门控阳离子通道**（voltage-gated cation channel），负责产生动作电位。动作电位的引发是通过质膜的去极化（depolarization）——即通过将膜电位降至更负（下面我们将会看到怎样通过神经递质的作用引起去极化）。在神经细胞和骨骼肌细胞中，引起充分去极化的刺激物使**电压门控 Na^+ 通道**（voltage-gated Na^+ channel）开放，允许少量 Na^+ 顺着其电化学梯度进入细胞。正电荷的内流使膜进一步去极化，从而开放更多的 Na^+ 通道，允许更多的 Na^+ 进入细胞，使膜更进一步去极化。这一过程以自放大的方式持续进行，直到膜局部区域的电位在几分之一毫秒内从静息值（约－70mV）迅速升至接近 Na^+ 的平衡电位（约＋50mV）（专题 11-2）。此时，当促使 Na^+ 流动的电化学净驱动力几乎为零时，如果通道的开放构象是稳定的，那么细胞将会达到一个新的静息状态，并伴随着所有 Na^+ 通道的持久开放。细胞通过协同作用的 2 种机制从这种持久的电痉挛中存活下来：Na^+ 通道的失活和电压门控 K^+ 通道的开放。

Na^+ 通道存在一种自动失活机制，即使膜仍被去极化，也能使通道快速重关闭。Na^+ 通道保持这种失活状态，不能重开放，直到膜电位回复至起始的负值之后数秒。因此，Na^+ 通道能以 3 种不同状态存在——关闭、开放和失活。图 11-27 显示它们如何促成动作电位的升降。

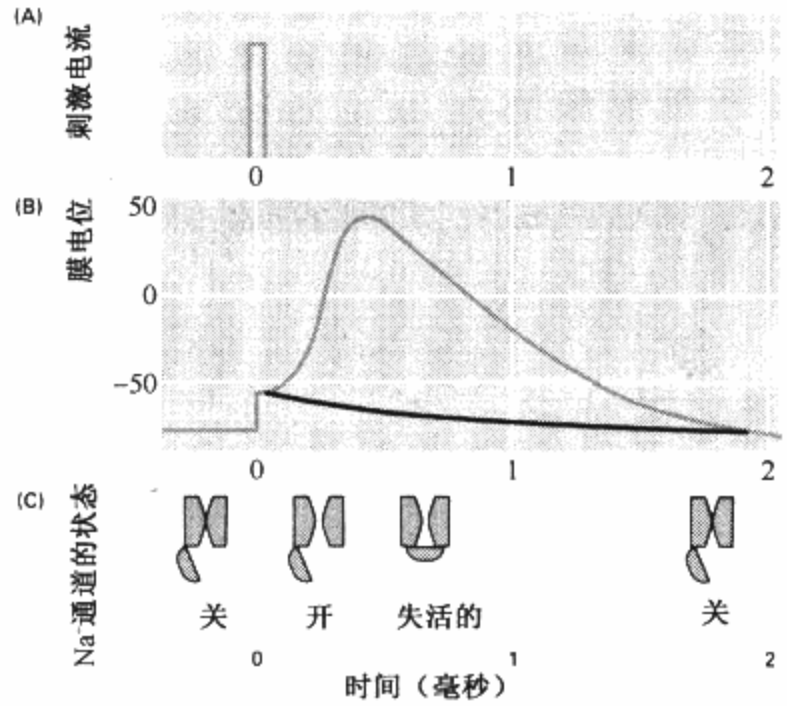


图 11-27　一个动作电位

（A）动作电位由一个短暂的电流脉冲触发。（B）使膜部分去极化，如动作单位对时间的作图所示。绿色曲线表示，如果膜中没有电压门控离子通道，膜电位在最初的去极化刺激之后怎样恢复；膜电位相对缓慢地回复至其－70mV 的初始值（没有开放的 Na^+ 通道）是因为 K^+ 通过 K^+ 通道外流，该通道为响应膜的去极化而开放，并驱动膜朝向 K^+ 平衡电位回复。红色曲线表示，电压门控 Na^+ 通道的开放及随后的失活所触发的动作电位的过程，该通道的状态如图中（C）所示。直到 Na^+ 通道返回其闭合构象，膜才能触发第二个动作电位；直到此时，膜才对刺激发生反应

刚才就动作电位的描述仅涉及一小块质膜。然而，这小块质膜的自放大去极化足以使质膜的邻近区域去极化，然后这些邻近区域经历相同的循环。如图 11-28 所示，动作电位通过这种方式从去极化的起始位点以行波形式进行传播，从而波及整个质膜。

在大多数神经细胞中，**电压门控 K^+ 通道**提供了第 2 种机制，有助于使活化的质膜更快回复至起始负电位，准备传递第 2 次冲动。这些通道开放的结果是，Na^+ 的瞬时内流被 K^+ 的外流迅速覆盖，使膜甚至在 Na^+ 通道完全失活之前很快回复至 K^+ 的平衡电位。这些 K^+ 通道对膜电位变化的应答方式与 Na^+ 通道几乎相同，仅在动力学上稍慢。由于这一原因，有时称它们为延迟 K^+ 通道（delayed K^+ channel）。

与 Na^+ 通道类似，电压门控 K^+ 通道可失活。电压门控 K^+ 通道突变体研究显示，通道蛋白 N 端的 20 个氨基酸对通道的快速失活是必需的。如果将该区域突变，那么通道失活的动力学就会发生改变，如果将该区域完全去掉，那么通道也就不会失活。令人惊异的是，在后一种情况下，如果将质膜的胞质面暴露于与丢失的氨基末端一致的合成小肽，那么可以恢复失活。这些发现暗示，每个 K^+ 通道亚单位的氨基末端的作用就像一个球，在孔道开放后不久即封闭其胞质端，从而使通道失活（图 11-29）。电压门控 Na^+ 通道的快速失活也有类似的机制（将在后面讨论），尽管似乎牵涉到该蛋白质的不同片段。

动作电位的电化学机制最初是通过 20 世纪 40 年代和 50 年代进行的一系列著名实验而被建立起来的。因为当时研究小细胞中的电事件的技术尚未建立，所以这些实验利用了乌贼的巨大神经元。尽管自那时起在技术方面取得了许多进展，但是最初应用的逻辑分析法作为现今研究工作的模型而被继续应用。专题 11-3 概括了最初的一些关键实验。

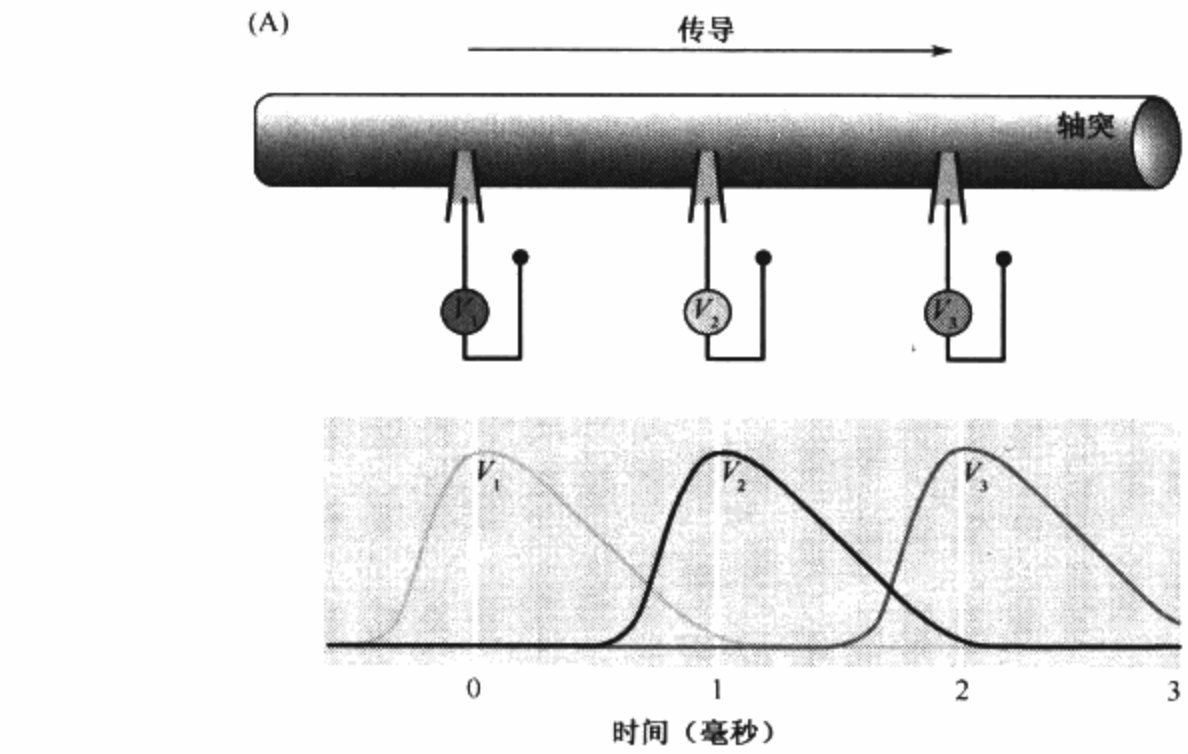


图 11-28　动作电位沿轴突的传播

（A）置于轴突之间的一组胞内电极所记录的电压。（B）Na^+ 通道和电流（橙色箭头）的变化引起膜电位的传播障碍。膜被去极化的轴突区域显示为蓝色底纹。注意，动作电位只能从去极化位点向远处传播，因为 Na^+ 通道的失活阻止去极化向后传播（图 11-30）。在有髓鞘的轴突上，Na^+ 通道簇彼此相距数毫米

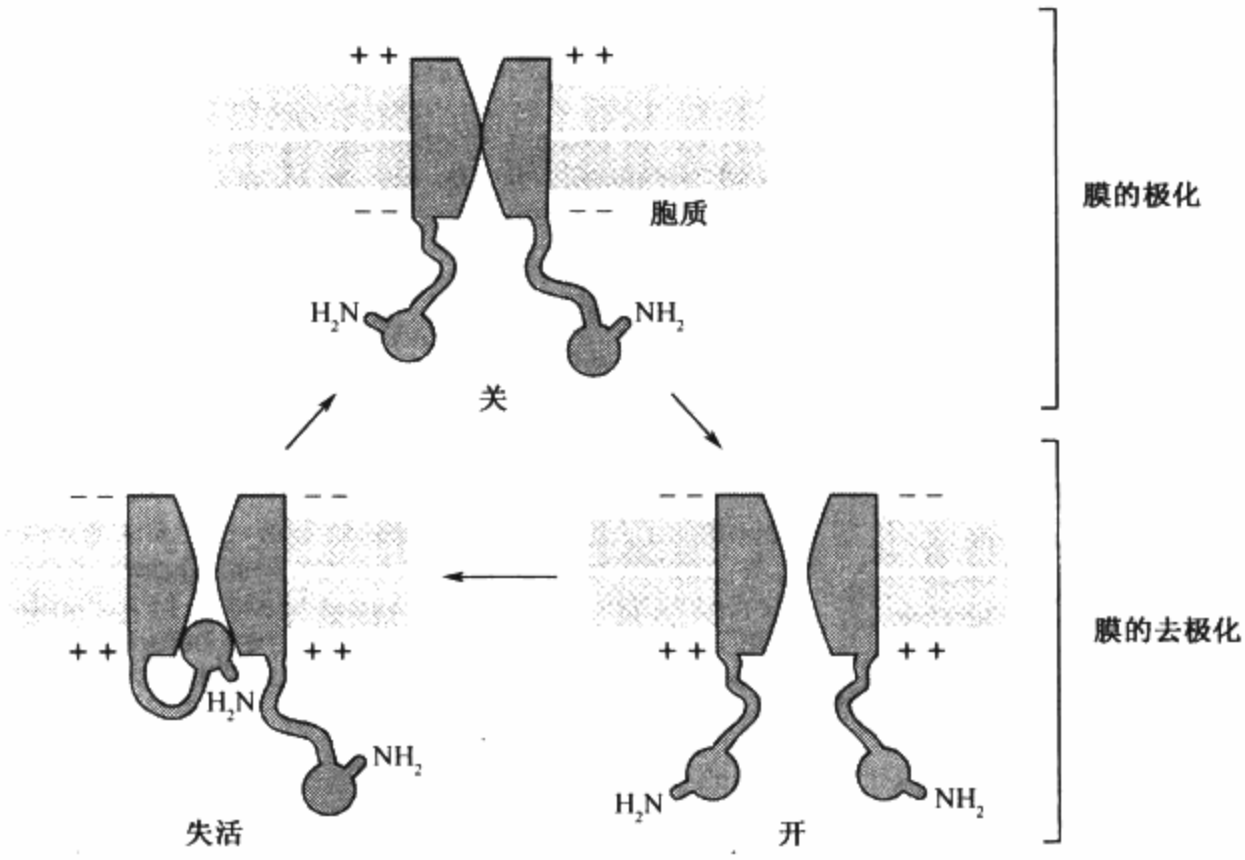


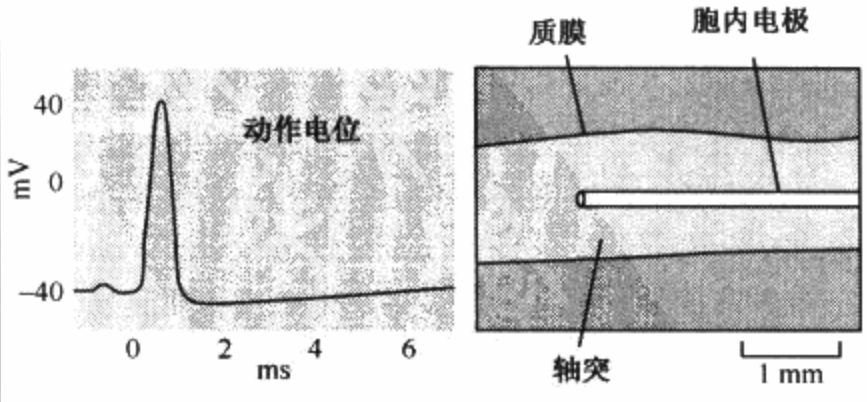
图 11-29　电压门控 K^+ 通道快速失活的“球-链”模型

当膜电位去极化时，通道开放，并开始传送离子。如果维持去极化状态，开放的通道就会采取一种失活构象，此时孔道就会被 N 端的 20 个氨基酸“球”占据，这个氨基酸“球”通过一段起“链”作用的未折叠多肽链与合适的通道相连。为简单起见，图中仅显示 2 个球；实际上有 4 个球，每个亚单位形成一个。一般认为，一种利用不同的多肽链片段的类似机制在 Na^+ 通道的失活中起作用。内部力量使每种状态在对抗微小紊乱中稳定下来，但是与其他分子发生十分强烈的碰撞也能使通道从这些状态中的一种转换到另一种。最低能量状态取决于膜电位，因为不同的构象有不同的电荷分布。当膜处于静息状态时（高度极化），闭合构象的自由能最低，因而最稳定；当膜被去极化时，开放构象的能量较低，因此通道开放的可能性很大。但是失活状态的自由能仍然较低；在开放状态的随机可变期之后，通道被失活。因此，开放构象符合仅能瞬时存在的相对稳定状态。红色箭头表示突然去极化之后相继发生的事件；黑色箭头表示膜被再极化后恢复至作为最低能量状态的最初构象

专题 11-3　用乌贼巨轴突所做的一些经典实验

1. 动作电位是用胞内电极记录的

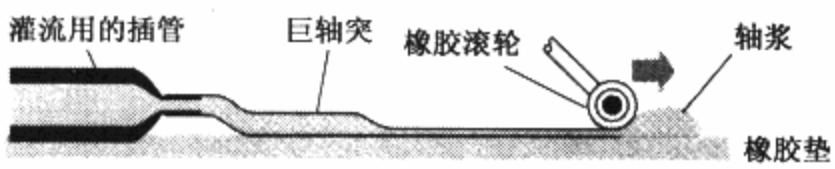
乌贼巨轴突的直径为 0.5～1mm，长为数厘米。电极的形状是一根含有导电溶液的玻璃毛细管，能沿着轴突的轴线刺入，以便其尖端能深入胞质。在胞内电极的帮助下，当动作电位经过电极时，人们能测量轴突内外的电位差，即膜电位。动作电位由作用于轴突一端的短暂电刺激触发。作用于哪一端并不重要，因为兴奋能向两个方向的任一一个方向传递；刺激有多大也不重要，只要它能超过特定的域值：动作电位是全或无的。



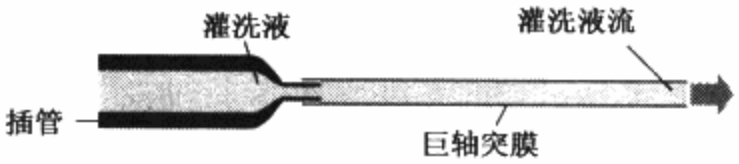
2. 动作电位仅依赖于神经元质膜和跨膜的 Na^+ 和 K^+ 梯度

轴突内外含量最丰富的 3 种离子是 Na^+、K^+、Cl^-。如在其他细胞中一样，Na^+-K^+ 泵维持浓度梯度：轴突内的 Na^+ 浓度约比轴突外低 9 倍，而轴突内的 K^+ 浓度约比轴突外高 20 倍。对动作电位而言，哪种离子重要呢？

乌贼巨轴突非常巨大、粗壮，以致像挤牙膏一样从轴突除去胞质是可能的，然后用 Na^+、K^+ 和 Cl^- 或 SO_4^{2-} 的纯净人工溶液从内部灌洗轴突。显然，如果（或只要）



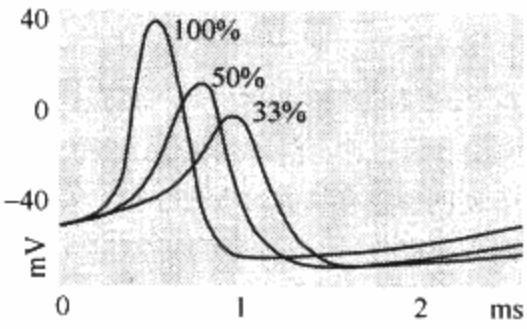
轴突内外的 Na^+ 和 K^+ 浓度接近生理状况，轴突仍将传播正常形式的动作电位。因此，细胞进行电信号传导的重要部分必定是质膜；重要的离子是 Na^+ 和 K^+；触发动作电位的充足的自由能来源必定是由跨膜浓度梯度来提供，因为其他所有来源的代谢能可能都已经通过灌流而被除去。



3. 在静息状态时，膜主要对 K^+ 有渗透性；在动作电位期间，它暂时对 Na^+ 有渗透性

在静息状态时，膜电位接近 K^+ 的平衡电位。当胞外 K^+ 浓度发生改变时，静息电位的变化大致与能斯特方程对 K^+ 的计算结果一致（专题 11-2）。因此，在静息状态时，膜主要对 K^+ 有渗透性：K^+ 漏电通道提供了一条通过膜的主要离子通道。

如果胞外的 Na^+ 浓度发生改变，那么对静息电位将不会有影响。然而，动作电位峰值高度的变化大致与能斯特方程对 Na^+ 的计算结果一致。因此，在动作电位期间，膜似乎主要对 Na^+ 有渗透性：Na^+ 通道开放。动作电位之后，膜电位回复至依赖于胞外 K^+ 浓度的负值，甚至比静息电位更接近 K^+ 的平衡电位：膜失去了对 Na^+ 的大部分渗透性，对 K^+ 的渗透性比以前大得多，即 Na^+ 通道关闭，额外的 K^+ 通道开放。

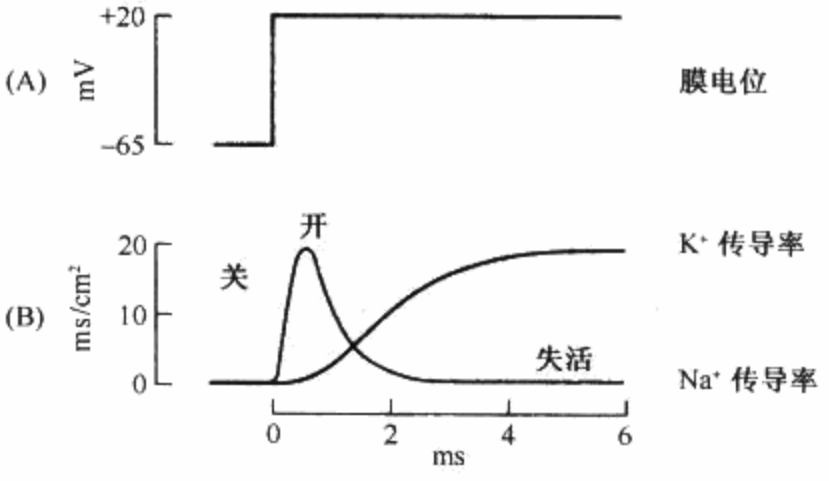


当胞外介质中含有正常 Na^+ 浓度的 100%、50%或 33%时形成的动作电位。

4. 电压钳制技术揭示膜电位怎样控制离子通道的开和关

在用另一个胞内电极监测膜电位时，通过使适当的电流流经一根沿轴突的轴插入的裸露金属丝，可使膜电位在整个轴突维持恒定（电压钳制）。当膜突然从静息电位转换并保持在去极化状态时（A），Na^+ 通道迅速开放，直到膜对 Na^+ 的渗透性远远大于对 K^+ 的渗透性；然后又同时关闭，即使膜电位被钳制且不发生变化。K^+ 通道也开放，但会有滞后，结果 K^+ 的渗透性上升，而 Na^+ 的渗透性下降（B）。如果此时通过将使膜迅速回复至静息电位，然后再次使膜快速去极化而非常迅速的重复这些实验，那么反应是不同的：持久去极化使 Na^+ 通道进入去极化状态，结果第二次去极化不能引起与第一次类似的膜电位的升降。从这种状态恢复需要相当长的时间（约 10ms）用于重建再极化（静息）膜电位。

在一个不受钳制的正常轴突中，Na^+ 通过开放的 Na^+ 通道涌入轴突会产生动作电位峰；Na^+ 通道的失活和 K^+ 通道的开放使膜快速回复至静息电位。



髓鞘形成增加动作电位在神经细胞中传播的速度和效率

许多脊椎动物神经元的轴突被一层**髓鞘绝缘**，这大大增加了轴突传播动作电位的速率。脱髓鞘病——多发性硬化（multiple sclerosis），戏剧性地证明了髓鞘形成的重要性。在这种疾病中，中枢神经系统某些区域中的髓鞘被破坏，神经冲动在这些区域的传播速度被大大减慢，常常产生破坏性的神经学后果。

髓鞘由被称为神经**胶质细胞**（glial cell）的专门支持细胞形成。**施旺细胞**（Schwann cell）在外周神经的轴突形成髓鞘，而**少突神经胶质细胞**（oligodendrocyte）则在中枢神经系统中发挥相同作用。这些神经胶质细胞以紧密螺旋形式围绕轴突一层一层地包裹自身质膜（图 11-30），从而绝缘轴突膜，以至几乎没有电流渗漏跨膜。这种髓鞘在间隔规则的郎飞结（node of Ranvier）处被中断，此处几乎集中了轴突中所有的 Na^+ 通道。因为轴突膜被包盖的部分有极好的电缆特性（换句话说，它们的电行为很像精心设计的水底电报电缆），所以膜在一个郎飞结处的去极化几乎立即被动传至另一个郎飞结。因此，动作电位沿着有髓鞘的轴突从一个郎飞结迅速传播至另一个郎飞结，这一过程被称为跳跃传导（saltatory conduction）。这种类型的传导主要有两个优势：动作电位传播更快，且节省代谢能，因为有效兴奋局限于郎飞结处轴突质膜的小区域。

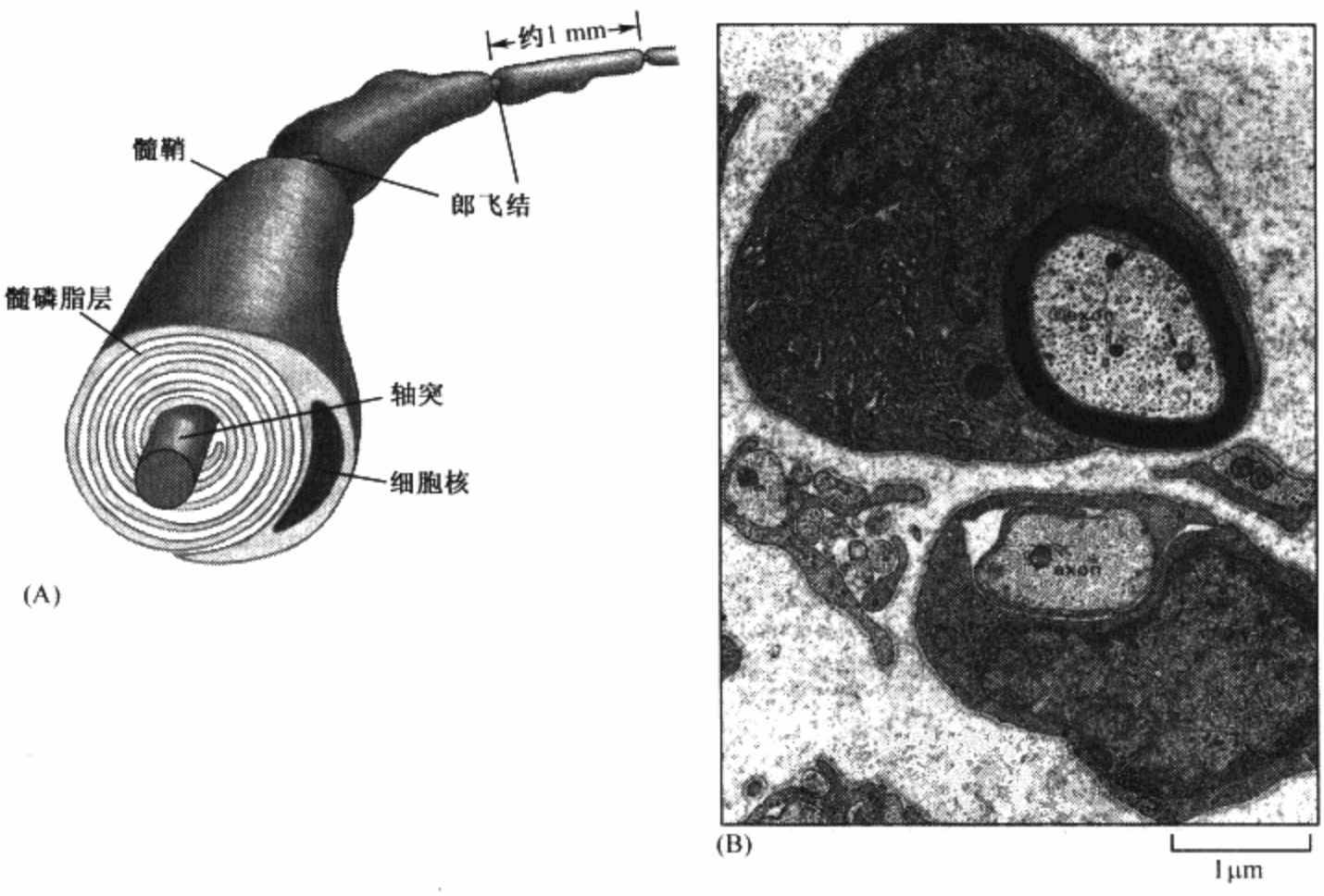


图 11-30　髓鞘形成

（A）一条来自外周神经的有髓鞘包围的轴突。每个许旺细胞将其质膜以同心圆的形式围绕轴突，形成一段长约 1mm 的髓鞘。为简洁起见，图中没有将髓鞘层显示成像在真实状态下一样紧密地结合在一起［图（B）］。（B）来自一只年轻大鼠腿部的一条神经的电子显微照片。图中可见到两个许旺细胞：靠近底部的一个细胞正开始在其轴突形成髓鞘；上面的一个细胞已经形成了一层近乎成熟的髓鞘［（B）摘自 Cedric S. Raine, in Myelin (P. Morell, ed.). New York: Plenum, 1976］

膜片钳记录表明单个门通道是以全或无的形式开放

神经元和骨骼肌细胞的质膜含有数千个电压门控 Na^+ 通道，跨膜电流是流经所有这些通道的电流的叠加。这种总电流可用胞内微电极记录，如图 11-28 所示。不过，记录流经单个通道的电流显然也是可能的。这可通过**膜片钳记录**的方式而获得，这是改革离子通道研究的一种方法，它允许研究人员检测通过在覆盖微量吸管嘴的一小片膜中的单分子通道蛋白的运输情况（图 11-31）。有了这种简单而有力的技术，就可以在所有类型的细胞中研究离子通道的详细性质。这项工作发现，即使没有电兴奋性的细胞，通常在其质膜中也有许多门控离子通道。这些细胞中的许多细胞，包括酵母，因太小而不能通过传统的生理学家用胞内微电极刺穿的方法来研究。

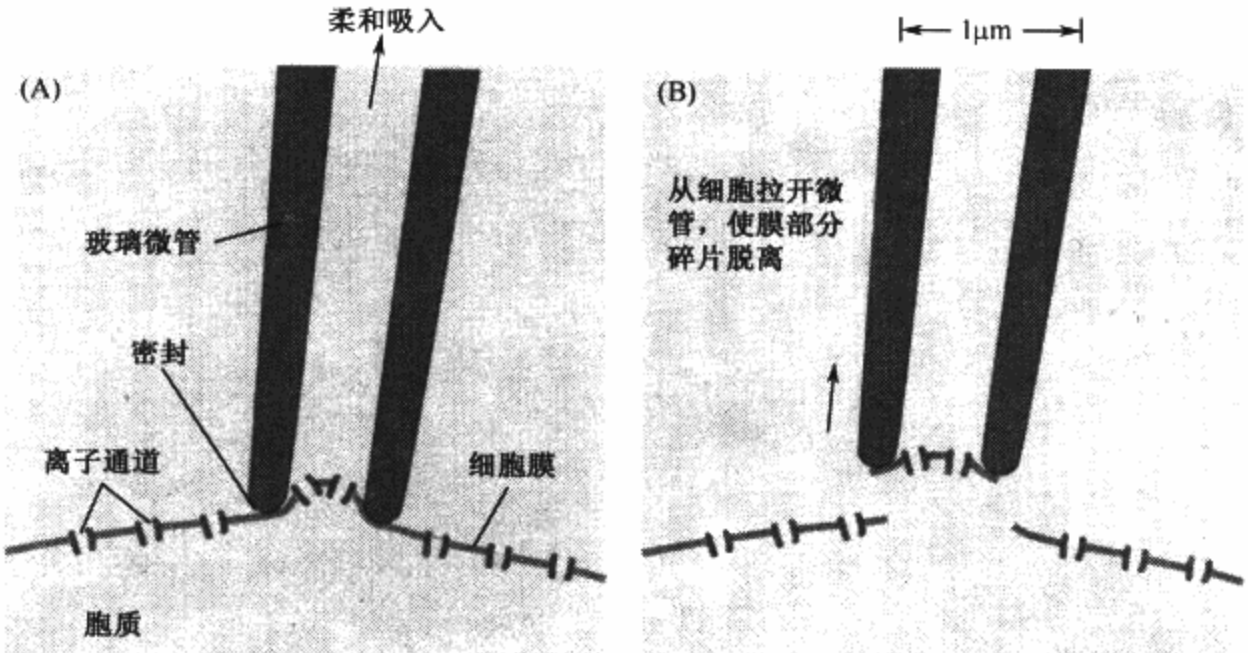


图 11-31　膜片钳记录技术

由于微量吸管和膜之间被极其紧密的密封在一起，所以电流仅通过流经覆盖其顶端的一小片膜中的通道就能进入或离开微量吸管。使用术语“钳”是因为在记录经过各个通道的离子电流时，将膜电位维持或“钳”在一个设定值要用到一种电子装置。通过这些通道的电流可被记录下来，那一小片膜仍然附着在细胞的其他部位，如（A）所示，或者脱落，如（B）所示。这一小片膜脱落的优点是，易于改变膜两侧的溶质成分，以便检测各种不同溶质对通道行为的影响。膜也可以相反的方向脱落，结果膜的胞质面朝向吸管的内侧

膜片钳记录表明，各个电压门控 Na^+ 通道以全或无的形式开放。通道开闭次数是随机的，但是，通道在开放时总是有相同的大传导率，每毫秒允许超过 1000 个离子通过。因此，一个完整细胞的跨膜总电流所表示的不是一个典型通道开放的程度，而是在任一时刻该细胞的膜上通道开放的总数（图 11-32）。

电压门控现象可以用简单的物理学原理来理解。静息神经元或肌肉细胞内部的电势比外部介质更低约 50～100mV。尽管这种电势差异似乎很小，但是它所跨的质膜厚度仅仅为大约 5 nm，结果，由此而产生的电压梯度约为 100 000V/cm。膜上的蛋白质因而处于一个很强的电场中。像所有其他的蛋白质一样，这些蛋白质有许多电荷基团，并在它们的各种不同原子之间存在许多极性键。因此这一电场将力强加于分子结构上。对许多膜蛋白而言，膜电场变化的影响很可能无关轻重的，但是电压门控离子通道能采取其稳定性依赖于电场强度的许多选择性构象。例如，电压门控 Na^+、K^+ 和 Ca^{2+} 通道在其中一个跨膜片段上存在特征性的正

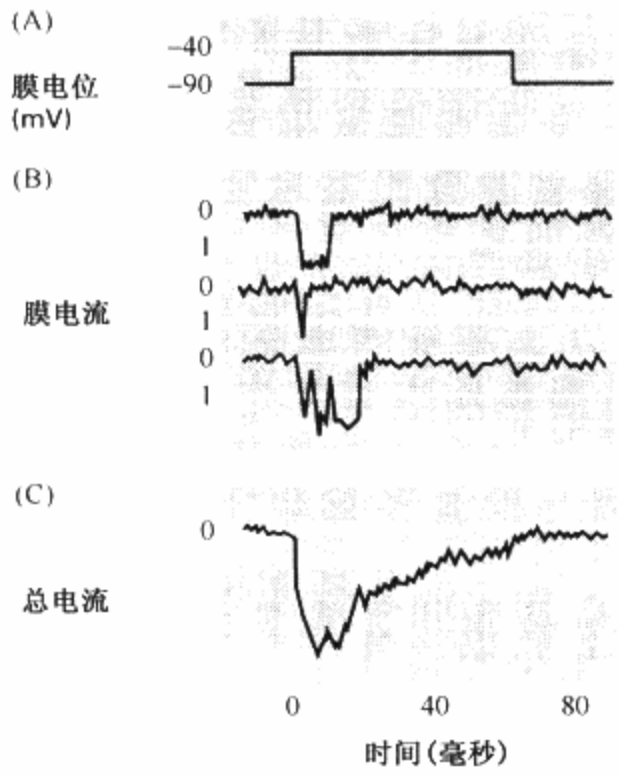


图 11-32　膜片钳测量单个电压门控 Na^+ 通道

如图 11-31 所示，从胚胎大鼠的肌肉细胞分离一片极小的质膜。（A）突然改变电位使膜去极化。（B）用同一片膜所作的 3 个实验的 3 条电流记录。（B）中每个主要的电流阶跃代表单个通道的开放和闭合。比较 3 条记录的结果显示，尽管通道开放和关闭的持续时间变化很大，但是经过一个开放通道的电流的比率实际上是恒定的。电流记录的微小波动主要来自记录仪器的电噪声。电流的测量结果以皮安（pA）表示。（C）同一实验重复测量电流 144 次的总和。这一整合电流相当于平常所见的 Na^+ 电流流经含有 144 个通道的一个相对较大区域。（B）和（C）之间的比较显示，整合电流的时程反映任何一个通道将处于开放状态的可能性；当去极化膜中的通道采取失活构象时，这种可能性随着时间的推移而降低（数据摘自 J. Patlak and R. Horn, *J. Gen. Physiol.* 79：333～351，1982. © The Rockefeller University Press）

电荷氨基酸，通过向外移动而对去极化作出反应，引发使通道开放的构象变化。如果环境的随机热运动使其充分颠簸，那么每种构象都能“翻转”至另一种构象，并且关闭、开放和失活构象对抗翻转的相对稳定性因膜电位的变化而被改变（图 11-29）。

电压门控阳离子通道在进化上和结构上是相关的

Na^+ 通道并不是唯一一种能产生动作电位的电压门控阳离子通道。例如，在一些肌肉细胞、卵细胞和内分泌细胞中的动作电位依赖于电压门控 Ca^{2+} 通道而不是 Na^+ 通道。而且，在一些没有正常电活性的细胞类型中还发现了功能未知的电压门控 Na^+、K^+ 或 Ca^{2+} 通道。

这 3 类通道都有大量的结构和功能多样性，它们是由多基因和来自相同基因产生的 RNA 转录本的选择性剪接产生。但是，已知的电压门控 Na^+、K^+ 和 Ca^{2+} 通道的氨基酸序列显示惊人的相似性。暗示它们都属于一个在进化和结构上相关联的蛋白质大家族，它们的许多设计原则都是相同的。单细胞的酿酒酵母含有编码电压门控 K^+ 通道的单个基因，然而，秀丽线虫的基因组却含有编码不同而相关的 K^+ 通道的 68 个基因。这种复杂性表明，甚至一个仅由 302 个神经元组成的简单神经系统也利用大量不同的离子通道来做出应答。

当人遗传有编码离子通道蛋白的突变基因时，可能会患许多种神经疾病、肌肉疾病、脑疾病或心脏病，这取决于该基因的表达部位。例如，在骨骼肌细胞中，编码电压门控 Na^+ 通道的基因的突变可引起肌强直（myotonia），这种病理状况是随意伸缩后的肌肉松弛大大延迟，从而引起痛苦的肌肉痉挛。在一些情况下，这是因为异常通道不能正常失活；结果，Na^+ 在动作电位结束后仍持续进入胞内，反复地再引发膜的去极化和肌肉收缩。与此类似，影响脑中 Na^+ 通道或 K^+ 通道的突变能引起癫痫（epilepsy），许多组神经细胞的同步过度触发会引起癫痫性发作（痫疯或发作）。

递质门控离子通道在化学突触处将化学信号转换成电信号

神经信号在称为**突触**（synapse）的特化接触位点从一个细胞传递给另一个细胞。通常的传递机制是间接的。细胞彼此之间是电隔离的，突触前细胞（presynaptic cell）通过一条狭窄的突触裂隙（synaptic cleft）与突触后细胞（postsynaptic cell）隔离。突触前细胞的电势变

化引发细胞释放被称为**神经递质**（neurotransmitter）的信号小分子，这种小分子被储藏在膜包围的突触泡中，并通过胞吐作用被释放。神经递质横跨突触裂隙快速扩散，并通过与递质门控离子通道（transmitter-gated ion channel）结合而在突触后细胞中引起电变化（图11-33）。神经递质在分泌后即被迅速除去：它或者被突触裂隙中的特殊酶破坏，或者被释放它的神经末端或周围的神经胶质细胞吸收。重吸收是通过许多种 Na^+ 依赖型的神经递质载体蛋白介导的；神经递质通过这种方式再循环，从而允许细胞跟上高释放速率。快速去除保证突触处信号转导在空间和时间上的精确性。它降低了神经递质影响邻近细胞的可能性，并在下一个神经递质脉冲被释放之前疏通突触裂隙，以便使重复、快速的信号传导事件的时限能准确地传达给突触后细胞。我们将会看到，通过这类化学突触（chemical synapse）的信号传导的通用性和适应性比通过电突触（electrical synapse）（讨论见第19章）处间隙连接的直接电耦联要大得多，神经元也利用这种电突触，但其范围要小得多。

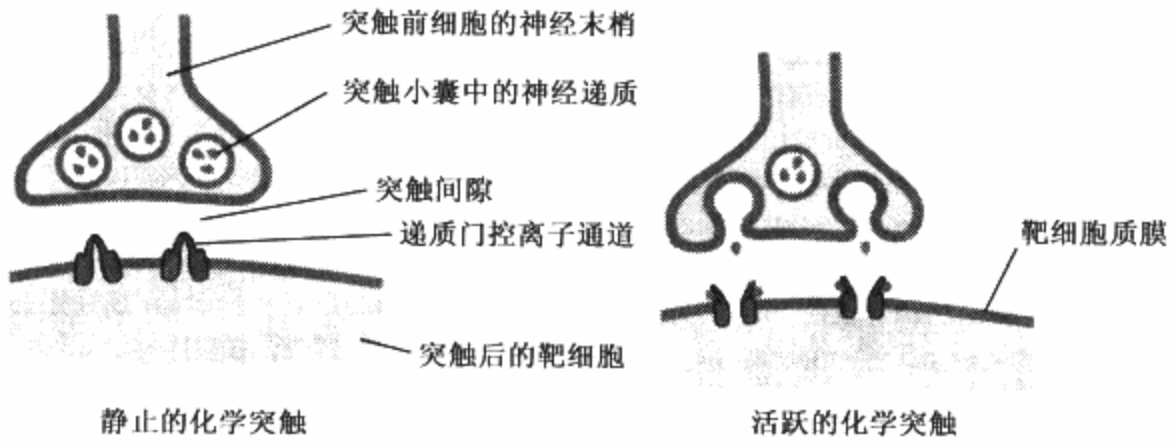


图 11-33　一个化学突触

当动作电位到达一个突触前细胞的神经末梢时，它刺激末梢释放神经递质。神经递质分子被包含在突触泡中，并且当突触泡与神经末梢的质膜融合时，被释放至胞外。释放的神经递质与递质门控离子通道结合并使之开放，这些通道集中于突触处的突触后靶细胞的质膜中。由此产生的离子流改变靶细胞的膜电位，从而传递受激神经发出的信号

递质门控离子通道专门用于将胞外化学信号在化学突触处快速转换成电信号。这些通道集中于突触区域内突触后细胞的质膜上，因应答神经递质分子的结合而瞬时开放，从而在膜中产生短暂的渗透性改变（图11-33）。与负责产生动作电位的电压门控通道不同，递质门控通道对膜电势相对不敏感，因此其本身不能产生自放大兴奋。相反，它们使局部渗透性发生改变，因而使膜电势发生改变，这些改变根据神经递质在突触处的释放量和持续时间的长短进行分级。只有当局部膜电位增加至足以使附近足够量的电压门控阳离子通道（存在于相同靶细胞膜上）开放，才可自这一位点引发动作电位。

化学突触可以是兴奋型的或是抑制性的

递质门控离子通道在几个重要方面彼此不同。首先，作为受体，它们有与突触前神经末梢释放的神经递质高度选择性结合的位点。第二，它们对其允许跨质膜通过的离子类型具有选择性；这决定了突触后应答的性质。**兴奋型神经递质**（excitatory neurotransmitter）使阳离子通道开放，引起 Na^+ 内流，从而使突触后膜朝触发动作电位的域值电位去极化。相反，**抑制性神经递质**（inhibitory neurotransmitter）或者开放 Cl^- 通道，或者开放 K^+ 通道，通过使兴奋性影响更加难以使突触后膜去极化而抑制触发动作电位。许多递质可能是兴奋性的，也可能是抑制性的，这取决于它们在何处被释放，它们与何种受体结合，以及它们遭遇到的

离子条件。例如，乙酰胆碱（acetylcholine）既有兴奋作用，又有抑制作用，这取决于与之结合的乙酰胆碱受体的类型。然而，通常乙酰胆碱、谷氨酸（glutamate）和5-羟色胺（serotonin）被用作兴奋性递质，而γ-氨基丁酸（γ-aminobutyric acid，GABA）和甘氨酸（glycine）则被用作抑制性递质。如谷氨酸在脊椎动物的脑中介导大多数兴奋性信号转导。

已经讨论了阳离子通道的开放如何使膜去极化。开放 Cl^- 通道的效应可理解如下。胞外 Cl^- 的浓度比胞内高得多（参见表11-1），但其内流受到膜电位的对抗。实际上，对许多神经元而言，Cl^- 的平衡电位接近静息电位——甚至更负。由于这一原因，所以 Cl^- 通道的开放会缓冲膜电位；由于膜开始去极化，所以更多带负电荷的 Cl^- 进入细胞对抗这一效应。因此，Cl^- 通道的开放使得膜的去极化更加困难，因而更加难以使细胞兴奋。K^+ 通道的开放具有类似效应。抑制性神经递质的重要性可通过阻断其作用的毒素的影响来说明。例如，番木鳖碱通过结合甘氨酸受体进而阻断甘氨酸的作用而引起肌肉痉挛、癫疯和死亡。

然而，并非神经系统中所有的化学信号都通过配基门控离子通道起作用。神经末梢分泌的许多信号分子，包括许多种神经肽，与仅间接调节离子通道的受体结合。这些所谓的G蛋白偶联受体（G-protein-linked receptor）和酶联受体（enzyme-linked receptor）将在第15章详细讨论。与递质门控离子通道结合的兴奋型和抑制性神经递质介导的信号转导通常是迅速、简单而短暂的，而与G蛋白偶联受体和酶联受体结合的配基介导的信号转导则倾向于缓慢和复杂得多，因而也更持久。

神经肌肉接头处的乙酰胆碱受体是递质门控阳离子通道

研究得最透彻的递质门控离子通道的例子是骨骼肌细胞的乙酰胆碱受体（acetylcholine receptor）。该通道通过**神经肌肉接头**处的神经末梢，即运动神经元和骨骼肌细胞之间的特化化学突触（图11-34），所释放的乙酰胆碱而被短暂开放。这种突触已被深入研究过，因为与中枢神经系统中的大多数突触不同，对这种突触进行电生理研究是很容易的。

乙酰胆碱受体在离子通道的研究历史中有特殊的意义。它是第一个被纯化的离子通道，其完整氨基酸序列最先被测定，在人造脂双层中最先被功能重建，其单个开放通道的电信号最先被记录。它的基因也是最先被克隆和测序的，它是唯一被测定其三维结构的配基门控通道，尽管仅有中等分辨率。纯化和鉴定这一受体快速进展的原因至少有两个：①异常丰富的乙酰胆碱受体源存在于电鱼和电鳐的电器官中（这些器官是经过改良的肌肉，发出强大的电休克用于捕食）。②某些蛇的毒液中的某些神经毒素（如α-银环蛇毒素）与该受体高亲和力（$K_a = 10^9\,L/mol$）和高特异性的结合，因而能通过亲和层析纯化该受体。荧光标记或放射性标记的α-银环蛇毒素也可用于定位和计数乙酰胆碱受体。通过这种方式显示，该受体被致密分布于神经肌肉接头处的肌肉细胞的质膜中（每平方微米约有20 000个这样的受体），但这些受体在同一质膜的其他部位相对很少。

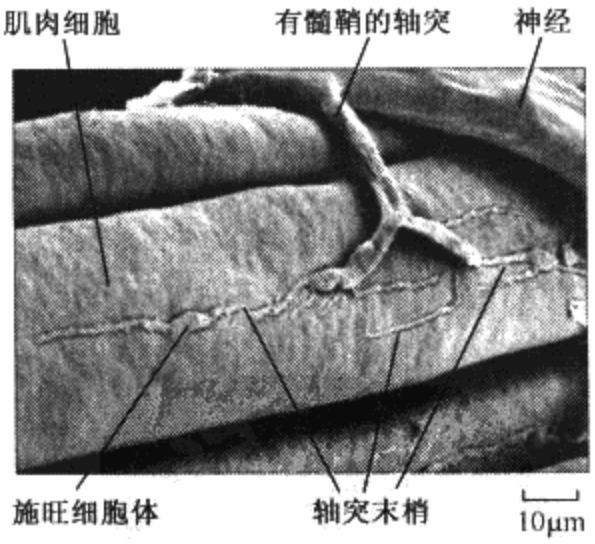


图11-34　青蛙的神经肌肉接点的低放大倍数扫描电子显微照片

图中显示骨骼肌细胞上单个轴突的末梢（摘自J. Desaki and Y. Uehara, *J. Neurocytol.* 10: 101～110, 1981. © Kluwer Academic Publishers）

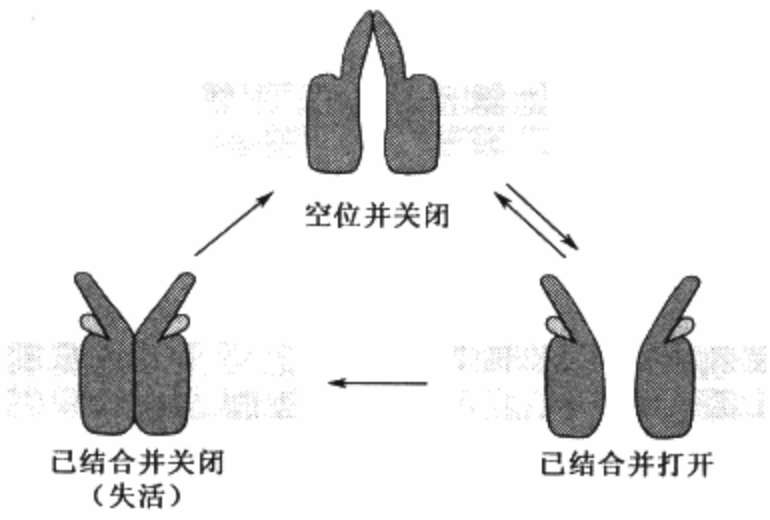


图 11-35　乙酰胆碱受体的 3 种构象

2 个乙酰胆碱分子的结合使递质门控离子通道开放。然后，它很可能保持开放状态，一直到乙酰胆碱被水解。然而，在乙酰胆碱持续存在的情况下，通道被失活（脱敏）。在正常情况下，乙酰胆碱被迅速水解，并且通道在大约 1ms 内，即正好在产生明显脱敏之前关闭。在乙酰胆碱持续存在大约 20ms 之后，将会出现脱敏现象

骨骼肌的乙酰胆碱受体由 5 条跨膜多肽组成，其中 2 条多肽相同，它们由 4 个独立的基因编码。这 4 个基因的序列极为相似，意味着它们是从同一个祖先基因进化而来。这一五聚体中 2 个完全相同的多肽每个都有乙酰胆碱的受体。当 2 个乙酰胆碱分子与五聚体复合体结合时，它们诱导复合体发生构象变化，使其通道开放。配基结合后，通道仍在开、放两种状态之间变换。但现在有 90%的可能性处于开放状态。这一状态一直持续到乙酰胆碱的浓度低至足以被定位于神经肌肉接头处的特异酶（乙酰胆碱酯酶）水解。一旦释放与之结合的神经递质，乙酰胆碱受体就恢复到它最初的静息状态。如果乙酰胆碱因过多的神经刺激而持续存在很长一段时间，那么该通道就会失活（图 11-35）。

乙酰胆碱受体的总体形状及其亚基的可能排列方式已通过电子显微镜技术测定（图 11-36）。5 个亚基排列成环状，形成一个由穿越脂双层的狭窄孔道组成的充满水的跨膜通道，在其两端扩宽形成孔腔。位于孔道的任何一个末端的荷负电氨基酸簇都有助于排斥阴离子和促进直径小于 0.65nm 的任何阳离子通过。正常运输主要由 Na^+ 和 K^+，还有一些 Ca^{2+} 组成。因此，与电压门控阳离子通道（如前面讨论过的 K^+ 通道）不同，它在阳离子之间几乎没有选择性，不同阳离子对通过通道的电流的相对贡献主要取决于它们的浓度和电化学驱动力。当肌细胞膜处于静息电位状态时，对 K^+ 的净驱动力接近零，因为电压梯度几乎平衡了跨膜 K^+ 浓度梯度（专题 11-2）。

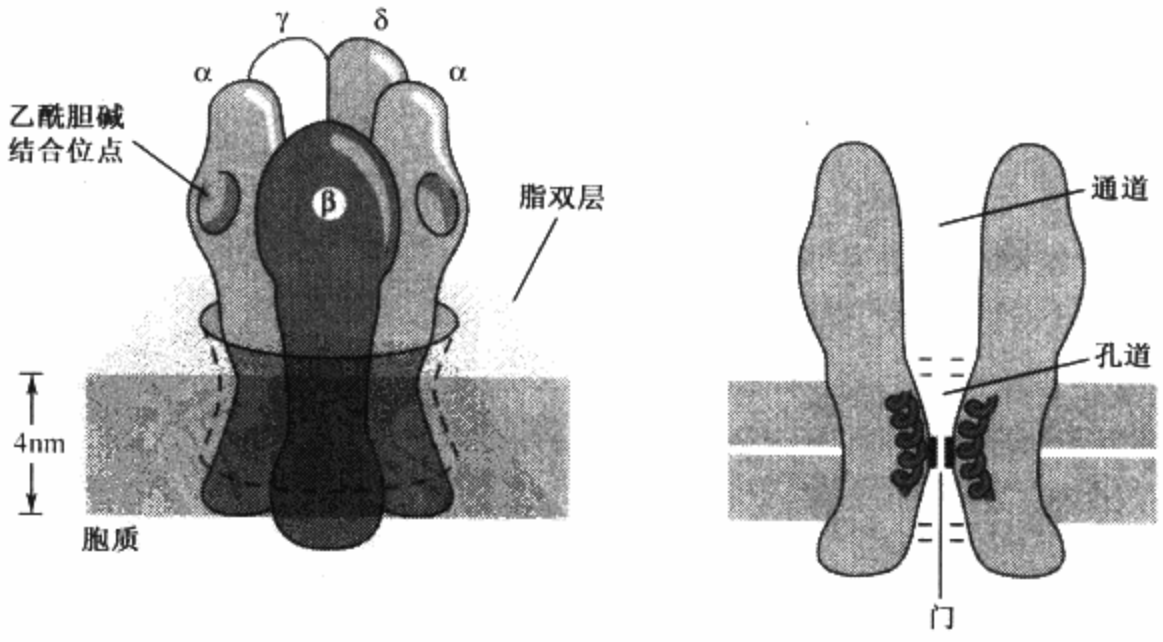


图 11-36　乙酰胆碱受体的结构模型

5 个同源亚基（α、α、β、γ、δ）结合，共同形成一个跨膜的含水孔道。每个亚基贡献一个 α 螺旋，5 个跨膜 α 螺旋沿孔道排列形成一个环状物，当孔道处于闭合构象时，一般认为它被 5 个亮氨酸的疏水侧链封闭，这些亮氨酸分别来自各个 α 螺旋，它们在靠近脂双层中央的位置形成一个闸门。在孔道的 2 个末端的荷负电侧链确保只有荷正电的离子通过孔道。2 个 α 亚基都含有乙酰胆碱结合位点；当乙酰胆碱与 2 个位点结合时，通道很可能通过使亮氨酸向外移动而发生构象变化，从而使闸门开放（改编自 N. Unwin, *Cell/Neuron* 72/10]：31～41，1993）

相反，对 Na^+ 而言，电压梯度和浓度梯度以相同方向发挥作用，驱动离子进入细胞（Ca^{2+} 的情况与此相同，但是 Ca^{2+} 的胞外浓度比 Na^+ 低得多，以致 Ca^{2+} 对总内向电流仅有很小的贡献）。因此，乙酰胆碱受体通道的开放导致了 Na^+ 的大量净内流（速率峰值为每个通道每毫秒约30 000个离子）。正如下面将要讨论的，这种内流引起膜的去极化，从而向肌肉发出收缩的信号。

递质门控离子通道是精神活性药物的主要靶点

应答神经递质乙酰胆碱、5-羟色胺、GABA 和甘氨酸而开放的离子通道含有结构上类似的亚基，暗示它们在进化上是有关联的，很可能以相同的方式形成跨膜孔道，即使它们的神经递质结合特异性和离子选择性明显不同。这些通道似乎有类似的总体结构，每一个由同源多肽亚基形成的通道很可能都装配成一个类似于乙酰胆碱受体的五聚体。谷氨酸门控离子通道由一个不同家族的亚基构成，一般认为是形成四聚体，这与前面讨论的 K^+ 通道类似。

对每类递质门控离子通道而言，每种亚基都存在两种形式，它们或者由不同的基因编码，或者由相同基因产物的选择性 RNA 剪接产生。这些形式综合不同的变异，用以形成一组极具多样性的不同通道亚型，具有不同的配基亲和力，不同的通道传导率，不同的开放和关闭速率，以及对药物和毒素的不同敏感性。例如，脊椎动物神经元的乙酰胆碱门控离子通道与肌细胞的不同，因为它们通常是由一种类型的 2 个亚基和另一种类型的 3 个亚基形成；但是它们至少有 9 个基因为第一种类型的亚基编码，为第二种类型的亚基编码的基因至少有 3 个，更多的多样性还可因 RNA 的选择性剪接引起。在脑中执行不同功能的乙酰胆碱敏感性神经元亚型的特征是这些亚基的不同组合。原则上以及实际在某种程度上，这使得设计靶向很窄的几组确定的神经元或突触的药物成为可能，从而特异性地影响特殊的脑功能。

实际上，递质门控离子通道长期以来都是重要的药物靶点。例如，外科医师在手术期间可通过使用箭毒（curare）——一种来源于植物的药物，最初被南美洲印第安人用于在箭上施毒，阻断骨骼肌细胞上的乙酰胆碱受体而使肌肉松弛。用于治疗失眠、焦虑、抑郁和精神分裂症的大多数药物在化学突触处发挥其效应，其中许多药物是通过与递质门控通道结合而发挥作用。例如，巴比妥和镇静剂，如地西泮（valium）和氯氮卓（librium），与 GABA 受体结合，通过允许较低浓度的这种神经递质开放 Cl^- 通道而使 GABA 发挥抑制作用。离子通道的新分子生物学通过揭示其结构的多样性和细节而使得有望设计新一代精神活性药物，这种药物将会更加选择性地减轻精神疾病的痛苦。

除离子通道之外，突出信号转导机制的许多其他组成成分都是精神活性药物的潜在靶点。正如前面所讨论的，许多神经递质在被 Na^+ 驱动的载体蛋白释放后，通过突触裂隙被清除掉。这种载体蛋白的抑制作用延长了递质的效应。从而促进突触传递。许多抗抑郁药，包括百忧解（prozac），都是通过抑制 5-羟色胺的吸收而发挥作用；其他药物则抑制 5-羟色胺和去甲肾上腺素的吸收。

离子通道是信号转导和估计的神经元装置得以构建的基本分子成分。为了了解这些装置的功能的可能复杂程度，我们考虑几个例子，用以说明几组离子通道在电兴奋细胞之间的突触通讯中怎样协同作用。

神经肌肉传递涉及 5 组不同离子通道的相继激活

离子通道对电兴奋细胞的重要性可通过神经冲动借以刺激肌细胞收缩的下述过程加以说明。这一似乎简单的应答需要至少 5 组不同的离子通道在几毫秒内被相继激活（图 11-37）。

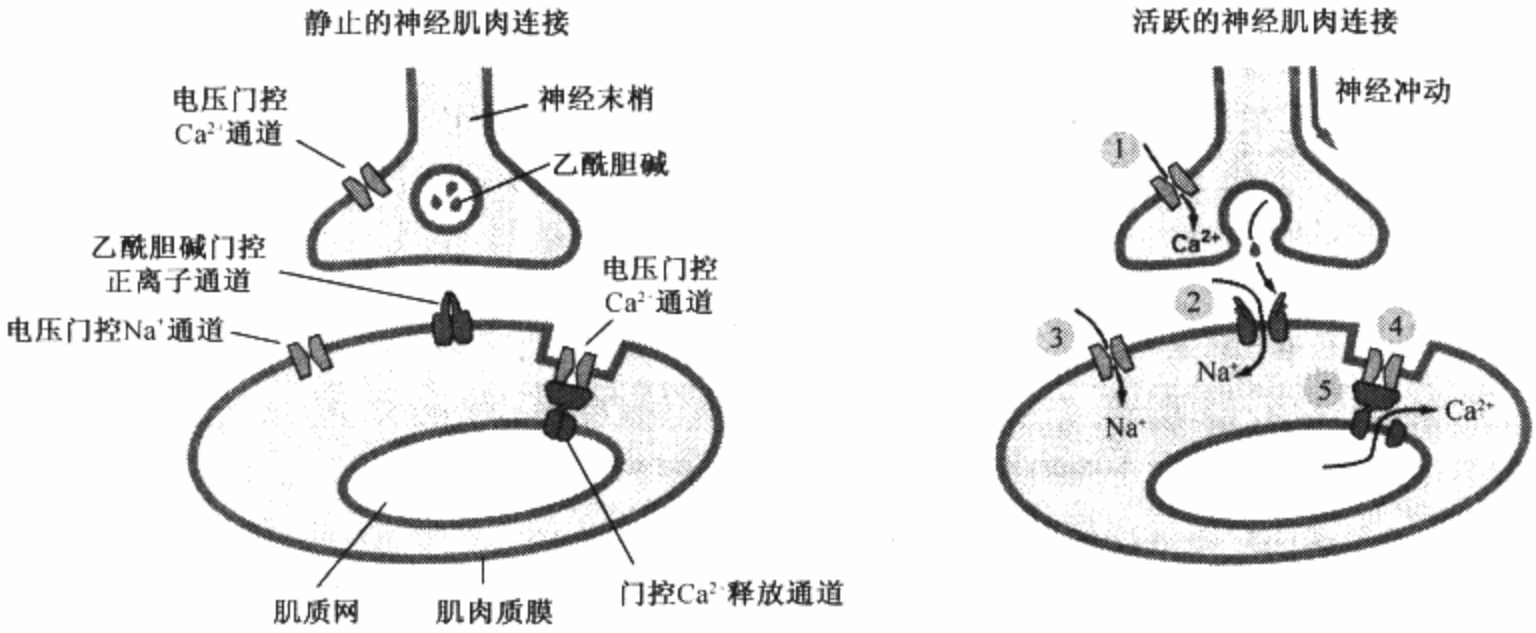


图 11-37　肌肉神经接点处的离子通道系统

这些门控离子通道对神经冲动刺激肌肉收缩是必不可少的。各种不同的通道以它们被激活的顺序（这已在文中描述）编号

1. 这一过程开始于神经冲动到达神经末梢并使末梢的质膜去极化时。去极化使膜中的电压门控 Ca^{2+} 通道短暂开放。由于胞外 Ca^{2+} 浓度比胞内游离 Ca^{2+} 浓度大 1000 倍，所以 Ca^{2+} 流入神经末梢。神经末梢胞质溶胶中 Ca^{2+} 浓度的增加触发乙酰胆碱局部释放进入突触间隙。

2. 释放的乙酰胆碱与肌细胞质膜中的乙酰胆碱受体结合，从而短暂开放与之相关的阳离子通道。由此导致的 Na^+ 内流使膜局部去极化。

3. 肌细胞质膜的局部去极化使膜中的电压门控 Na^+ 通道开放，从而允许更多的 Na^+ 进入胞内，并进一步使膜去极化。这依次使邻近的电压门控 Na^+ 通道开放，并导致去极化的自传播（动作电位），最终导致整个质膜的去极化膜（图 11-28）。

4. 肌细胞质膜的广泛去极化激活膜中特化区域（横向 T 小管——讨论见第 16 章）的电压门控 Ca^{2+} 通道。

5. 这依次使肌浆网膜邻近区域的 Ca^{2+} 释放通道（Ca^{2+} release channel）短暂开放，并将储藏在肌浆网中的 Ca^{2+} 释放进入胞质溶胶。胞质溶胶中 Ca^{2+} 浓度的突然增加使肌细胞中的肌原纤维收缩。还不能确定 T 小管膜中电压门控离子通道的激活怎样导致肌浆网膜中 Ca^{2+} 释放通道的开放。然而，这两种膜紧密贴在一起，这两种类型的通道通过一个特化结构结合在一起（图 16-74）。因此，电压诱导的质膜 Ca^{2+} 通道构象的变化通过一个机械连接器直接使肌浆网中的 Ca^{2+} 释放通道开放。

尽管运动神经元激活肌收缩是复杂的。但是，一个神经元要整合突触处的大量输入信号并计算出适当的输出，就需要离子通道之间更复杂的相互作用，现在就这一点予以讨论。

各个神经元都是一个复杂计算装置

在中枢神经系统中，一个单一的神经元能从数千个其他神经元接收输入信号，并能依次与数千个其他细胞形成突触。例如，数千个神经末梢与脊髓中的一般运动神经元形成突触；它的细胞体和树突几乎完全为它们所覆盖（图 11-38）。其中一些突触传递来自脑和脊髓的信号；其他突触则携带来自肌肉或皮肤的感觉信息。运动神经元必须综合来自所有这些来源的信息，并通过沿其轴突释放动作电位或通过保持沉默而做出应答。

在一个神经元的许多突触中，一些倾向于使神经元兴奋，而另一些则抑制神经元。兴奋

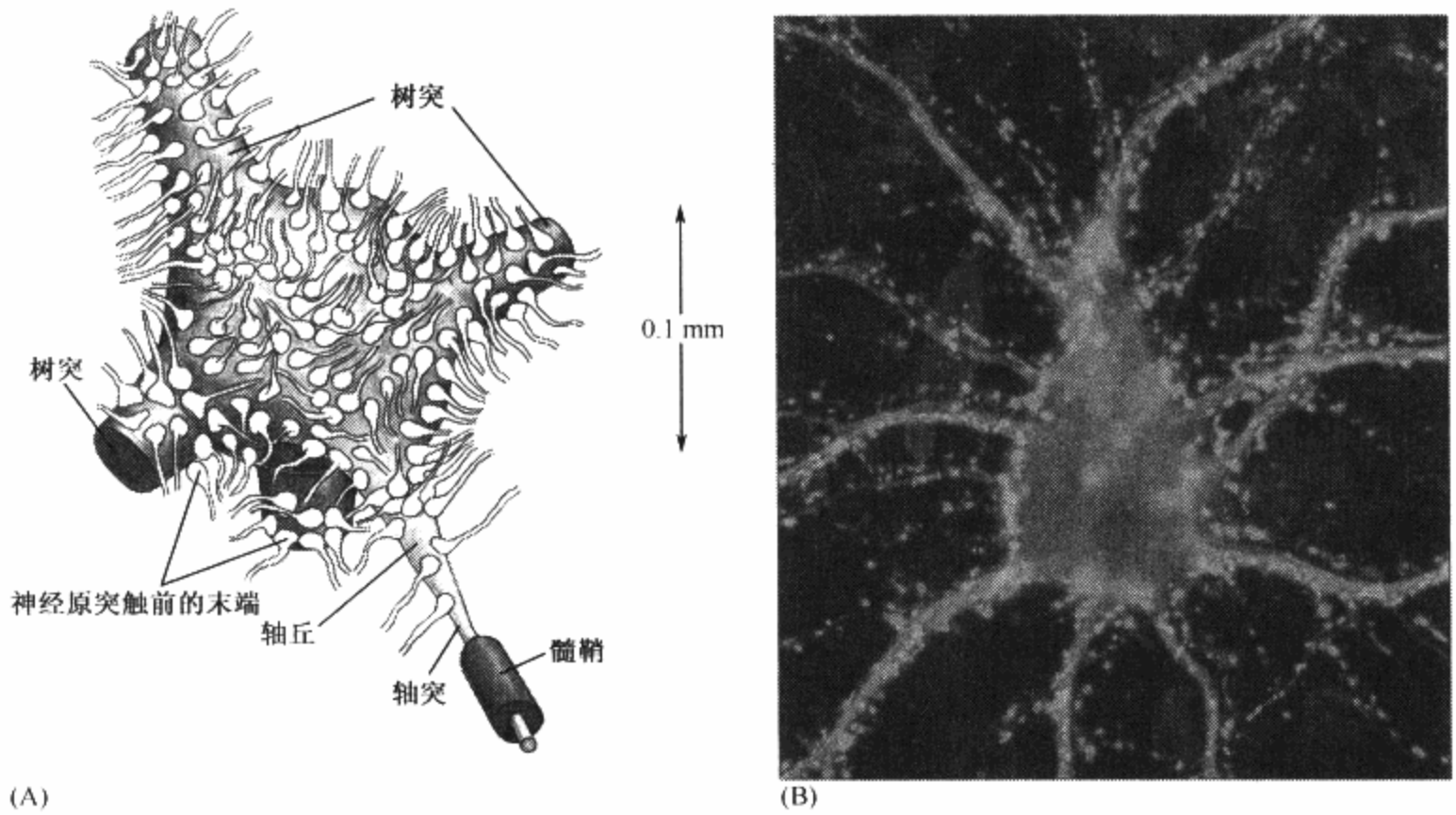


图 11-38　脊髓中的一个运动神经元细胞体

（A）在细胞体和树突上的数千种神经末梢突触。这些突触传递来自机体其他部分的信号，用以控制沿着这一巨大细胞的单一轴突触发动作电位。（B）显示被识别细胞骨架蛋白（绿色）的荧光抗体染色的神经细胞体和树突的显微照片。来自其他神经细胞（不可见）的数千个轴突末梢（红色）在细胞体和树突上形成突触；它们被识别突触泡中的一种蛋白质的荧光抗体染色［（B）经 Olaf Mundigl 和 Pietro de Camilli 许可］

性突触释放的神经递质使突触后膜轻微去极化，称为兴奋性突触后电位（兴奋性 PSP），而抑制性突触释放的神经递质通常引起轻微超极化，称为抑制性 PSP。大多数神经元的细胞体和树突的膜含有相对低密度的电压门控 Na^+ 通道，单个兴奋性 PSP 通常太小而不能引发一个动作电位。相反，每个输入信号表现为不同等级强度的局部 PSP，它随着距突触位点的距离的增加而下降。如果信号同时到达位于树突树相同区域中的几个突触，那么邻近区域的总 PSP 将大致是各个 PSP 的叠加，而抑制性 PSP 对总 PSP 的贡献是负向的。来自各个邻近区域的 PSP 被动扩散，并汇集于细胞体。因为与树突树相比，细胞体很小，所以它的膜电位大致相同，是作用于细胞上的所有信号的效应的组合，是根据突触距细胞体的距离进行加权的结果。因此，细胞体的组合 PSP 代表接收到的所有刺激的**空间叠加**（spatial summation）。如果兴奋性信号输入占优势，那么组合 PSP 就去极化；如果抑制性信号输入占优势，那么它通常就是超极化。

空间叠加组合了在不同膜位点接收到的信号的效应，而**时间叠加**（temporal summation）则组合了在不同时间所接收到的信号的效应。如果一个动作电位在突触的前 PSP 完全衰退之前到达突触，并引发神经递质的释放，那么第二个 PSP 加至存留下来的第一个 PSP 的后面。如果许多动作电位迅速相继到达，那么每个 PSP 就会加至前面的 PSP 的后面，叠加成一个大而持久的平均 PSP，其强度反映了突触前神经元的发放速率（图 11-39）。时间叠加的本质是：它将输入信号的频率翻译成一个净 PSP 的强度。

时间叠加和空间叠加共同提供了许多突触前神经元的发放速率联合控制单个突触后细胞体膜电位的手段。突触后细胞在神经计算中所做的最后一步是产生输出，这种输出通常以动作电位的形式，将一个信号转而传递给其他细胞。输出信号反映了细胞体中 PSP 的强度。尽

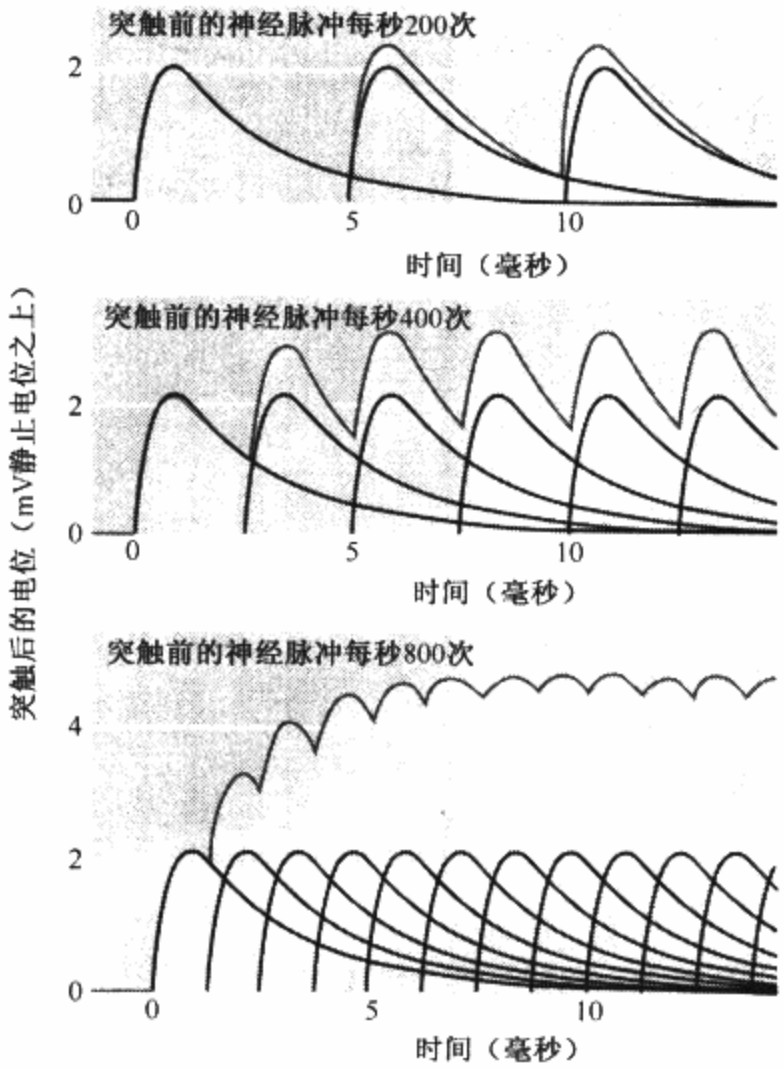


图 11-39　时间叠加的原理

到达一个突触的每个突触前动作电位都产生一个小突触后电位，或 PSP（黑线）。当动作电位相继到达同一突触时，产生的每个 PSP 都附加在前一个 PSP 的尾部，从而产生一个更大的组合 PSP（绿线）。新来的动作电位的频率越大，组合 PSP 也就越大

管 PSP 是连续分级的变量，然而，动作电位总是以全或无的形式存在，其大小是统一的。通过动作电位的信号转导中的唯一变量是一个动作电位和下一个动作电位之间的时间间隔。因此，对于长距离传递而言，PSP 的强度被翻译或被编码成动作电位的发放频率（图 11-40）。这种编码通过一组特殊的门控离子通道而获得，这些通道以高密度出现在邻近细胞体的突触基部，一个被称为轴丘的区域（图 11-38）。

神经计算至少需要 3 类 K^+ 通道的组合

我们已经看到，对远距离传递而言，神经元接收到的刺激强度由神经元触发的动作电位的频率决定：刺激越强，动作电位的频率就越高。动作电位起始于轴丘，即每个神经元的一个独特区域，含有大量的电压门控 Na^+ 通道。但是，为了执行其特殊的编码功能，**轴丘膜**也至少含有 4 种其他类型的离子通道——其中 3 个通道对 K^+ 具有选择性，一个通道对 Ca^+ 具有选择性。K^+ 通道的这 3 个变异体有不同的特性；我们将它们称为延迟 K^+ 通道、早期 K^+ 通道和 Ca^{2+} 激活的 K^+ 通道。

为了理解这种对多种类型通道的需求，首先考虑，如果神经细胞中存在的唯一电压门控离子通道是 Na^+ 通道，那么会发生什么情况。在低于突触刺激的特定域值水平时，轴丘膜的去极化将不足以触发一个动作电位。随着刺激的逐渐增加，去极化将达到这一域值，Na^+ 通道将会开放，最终将触发动作电位。动作电位将通过 Na^+ 通道的失活以通常的方式被终止。在另一个动作电位被触发前，这些通道必须从其失活状态中恢复过来。但是，这将要求膜电

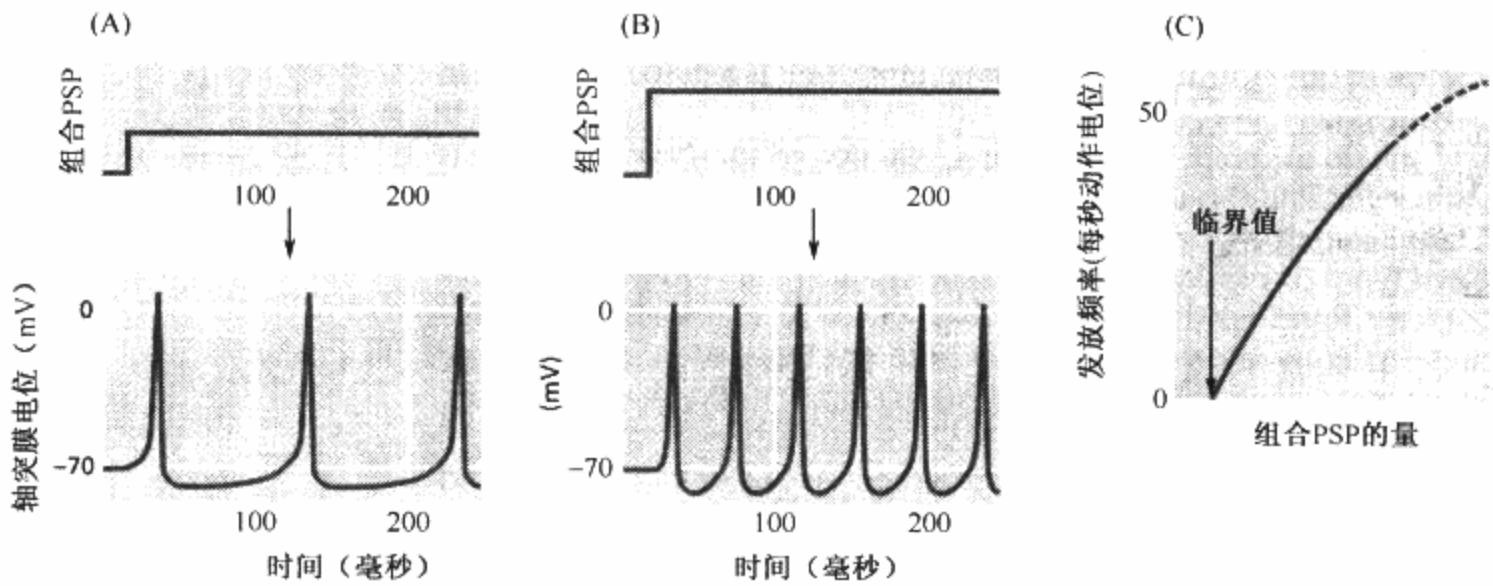


图 11-40　以轴突发放动作电位的频率的形式编码组合 PSP

（A）和（B）的比较显示轴突的发放频率怎样随组合 PSP 的增加而增加。（C）概括了它们之间的总体联系

压回复至一个非常负的值，只要强烈的去极化刺激（来自 PSP）继续存在，就不会发生这种回复。因此，需要另一种类型的通道使经历每个动作电位之后的膜再极化，以使细胞再次触发动作电位。

这一任务由前面讨论过的、与动作电位的传播有关的**延迟 K^+ 通道**来执行。它们是电压门控通道，但是由于它们的动力学较慢，所以它们仅在动作电位的下降阶段才开放，此时 Na^+ 通道处于失活状态。它们的开放允许 K^+ 外流，驱动膜回复至 K^+ 平衡电位，这一电位非常负，以致 Na^+ 通道迅速从其失活状态恢复。膜的再极化也使延迟 K^+ 通道关闭。此时轴丘复位，以便来自突触输入的去极化刺激能触发另一个动作电位。通过这种方式，树突和细胞体的持续刺激导致突触反复触发动作电位。

然而，反复触发实质上是不够的。触发的频率必须反映刺激的强度，一个由 Na^+ 通道和 K^+ 通道组成的简单系统不足以达到这一目的。细胞处于平稳刺激的特定域值水平以下时，根本不会触发动作电位；当高于这一域值水平时，它将以相对快的速率突然开始触发动作电位。**早期 K^+ 通道**可以解决这一问题。它们也是电压门控通道，当膜被去极化时开放，但是它们的特殊电压敏感性和失活动力学是：它们只降低正好高于触发所需域值的刺激水平的发放速率。这样，它们就消除了在发放速率和刺激强度之间的关系中的不连接性。其结果是，发放速率与去极化刺激的强度在很宽范围内成比例（图 11-40）。

这一编码过程通常会被开始时提到的轴丘中的另外两种离子通道（电压门控 Ca^{2+} 通道和 Ca^{2+} 激活的 K^+ 通道）做进一步调整。它们共同作用，降低细胞的应答至一个持续很长时间的不变刺激，此过程称为**适应**（adaptation）。这些 Ca^{2+} 通道类似于介导神经递质从突触前的轴突终末释放的 Ca^{2+} 通道；当动作电位被触发时，它们开放，暂时允许 Ca^{2+} 进入轴丘。

Ca^{2+} 激活的 K^+ 通道在结构上和功能上都不同于前面描述过的任何一种通道类型。它的开放是为了响应神经细胞膜胞质面增高的 Ca^{2+} 浓度。如一个强去极化刺激持续很长一段时间，引发一长串动作电位。每个动作电位允许 Ca^{2+} 通过电压门控 Ca^{2+} 通道的短暂内流，结果胞内 Ca^{2+} 浓度逐渐积累至足以使 Ca^{2+} 激活的 K^+ 通道开放的高水平。因此而增加的膜对 K^+ 的通透性使膜更加难以去极化，所以它增加了在一个动作电位和下一个动作电位之间的时间延迟。一个长期接受持续刺激的神经元通过这种方式对持久不变的刺激的反应性逐渐变得更低。

这种适应性（也可通过其他机制发生）允许神经元（通常是神经系统）对变化做出灵敏

的反应，甚至对高背景水平的稳定刺激。这是帮助我们感觉到，如肩膀上的轻轻接触，还有忽视衣服对我们的持久压力的策略之一。我们将在第 15 章以细胞信号传导过程中的一般特征来更加详细的讨论适应性。

其他神经元进行不同的计算，从而通过无数种方式对它们的突触输入做出应答，这反映了位于膜中的不同离子通道家族成员的不同分类。例如，在脊椎动物神经系统中至少有 5 种已知类型的电压门控 Ca^{2+} 通道和至少 4 种已知类型的电压门控 K^+ 通道。基因的多样性显然考虑到了宿主的不同类型神经元，其电行为与它们必须执行的特殊任务尤其一致。

神经系统的关键特性之一是它的学习和记忆能力，这似乎主要取决于特殊突触的长期变化。在本章最后，讨论一种异常类型的离子通道，一般认为它在某种形式的学习和记忆中有特殊的作用。它定位于中枢神经系统的许多突触中，这些突触由电压和兴奋性神经递质谷氨酸控制。它也是精神活性药物苯环利定或安琪儿粉的作用部位。

哺乳动物海马中的长程增强效应（LTP）依赖于 Ca^{2+} 通过 NMDA 受体通道的输入

实际上，所有动物都能学习，但是哺乳动物似乎特别善于学习（或者我们愿意这么认为）。在哺乳动物的大脑中，被称为海马（hippocampus）的区域在学习中具有特别的作用。当大脑两侧受到破坏时，形成新的记忆的能力大部分都丧失了，尽管以前的长期记忆仍然保留。相应地，海马中的一些突触因反复使用而表现出明显的功能性改变：在突触前细胞中偶尔出现的单个动作电位不会留下持久的痕迹，而反复触发则会引起**长程增强效应**（long-term potentiation，LTP），因为突触前细胞中随后到来的单个动作电位在突触后细胞中引起的应答大大增强。依据反复触发的数目和强度的不同，这一效应可持续数小时、数天或数周。只有被激活的突触才表现出 LTP；在同一突触后细胞保持静息状态的则不受影响。然而，当细胞正通过一组突触接受反复刺激时，如果单个动作电位在其表面的另一个突触处被传递，那么后者也将经历 LTP，尽管在另一时间传递的单个动作电位不会留下如此持久的痕迹。

这类突触的根本规则似乎是：每当突触后膜被强烈去极化时（通过同一突触前细胞的新近反复触发，或者通过其他方式），在突出前细胞触发（一次或多次）的任何时候都可发生 LTP。这一规则反映了突触后膜中一类特殊离子通道的行为。谷氨酸是哺乳动物中枢神经系统中主要的兴奋性神经递质，谷氨酸门控离子通道在脑的所有递质门控通道中是最常见的。像在其他部位一样，在海马中，造成兴奋性 PSP 的大部分去极化电流由以标准方式运行的谷氨酸门控离子通道来保持。但是，该电流还有一个更有趣的成分，它由被称为 **NMDA 受体**的谷氨酸门控离子通道的一个独特亚类介导，如此命名是因为它们被谷氨酸的人造类似物 *N*-甲基-D-天冬氨酸选择性激活。NMDA 受体通道是双重门控的，只有同时满足如下 2 个条件时才开放：谷氨酸必须与受体结合，膜必须被强烈去极化。第二个条件是释放 Mg^{2+}（在正常情况下封闭静息通道）所必需的。这意味着，NMDA 受体通常只在常规谷氨酸门控离子通道也被激活并使膜去极化时才被激活。NMDA 受体对 LTP 是极为重要的。当它们被特殊抑制剂选择性封闭时，LTP 不会发生，即使寻常的突触传递仍在继续。用这种抑制剂处理的动物表现出特殊的学习能力缺陷，但是另一方面，其行为几乎与正常无异。

NMDA 受体怎样介导这种惊人的效应呢？答案是，这些通道在开放时对在突触后细胞中作为胞内介质的 Ca^{2+} 具有高度渗透性，从而触发导致 LTP 的一连串变化。因此，当通过注射 Ca^{2+} 螯合剂 EGTA 而人为地保持突触后细胞内的低 Ca^{2+} 水平时，LTP 被阻止。增加突触后细胞对谷氨酸的敏感性的长期变化是在质膜中插入新的常规谷氨酸受体（图 11-41）。证据

也表明，变化还可发生于突触前细胞，结果，当细胞随后被激活时，会释放比平常更多的谷氨酸。

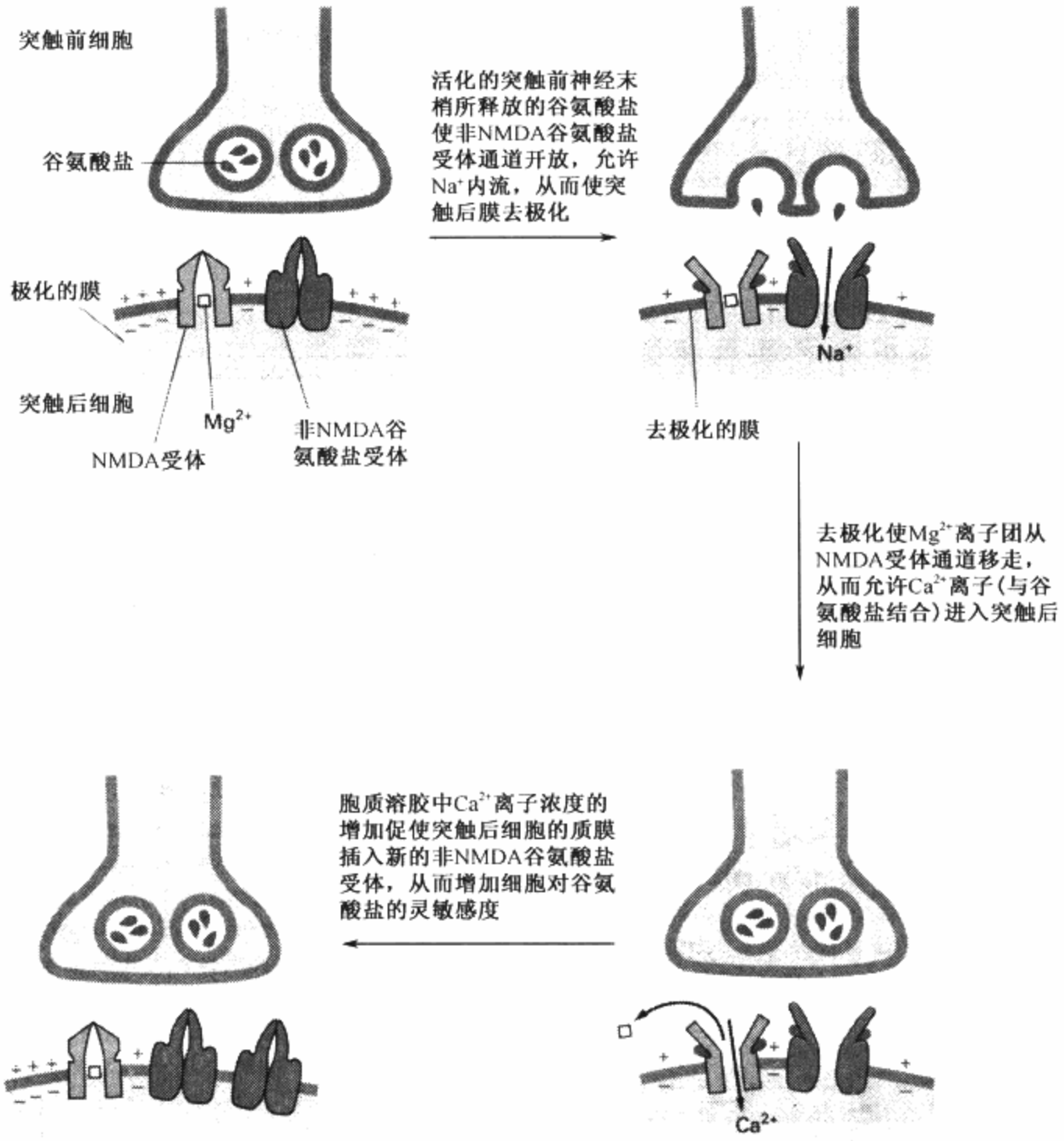


图 11-41　长时程增强过程中的信号转导事件

尽管图中没有显示，但是有证据表明，在 LTP 中，变化也可发生于突出前神经末梢，这可能是受到来自突触后细胞的逆行信号的刺激

证据表明，除海马之外，NMDA 受体在脑的其他部分的学习和相关现象中也起重要作用。而且，在第 21 章中将会看到，NMDA 受体在根据神经系统发育期间的经历调整突触连接的解剖学模式中有极为重要的作用。

因此，突触处释放的神经递质除中继传递瞬时电信号外，也能改变胞内介质的浓度，引起突触传递效率的持久改变。然而，仍不确定的是，在细胞成分正常周转的情况下，这些改变怎样持续数周、数月或一生。

表 11-2 总结了已经讨论过的离子通道家族。

表 11-2　一些离子通道家族

家族*	代表性亚家族	
电压门控阳离子通道	电压门控 Na^+ 通道	
	电压门控 K^+ 通道（包括延迟型和早期型）	
	电压门控 Ca^{2+} 通道	
递质门控离子通道	乙酰胆碱门控阳离子通道	兴奋性
	谷氨酸门控 Ca^{2+} 通道	兴奋性
	5-羟色胺门控阳离子通道	兴奋性
	GABA 门控 Cl^- 通道	抑制性
	甘氨酸门控 Cl^- 通道	抑制性

* 一个家族各个成员的氨基酸序列类似，因此认为它们来自共同的祖先；在亚家族内部，通常这种相似性更为接近。

小结

离子通道形成跨越质膜的含水孔道，允许适当大小和电荷的无机离子顺着其电化学梯度跨膜，其速率比通过任何已知运载体的速率约大 1000 倍。这些通道是“门控的”，通常短暂开放，用以响应膜中的特殊干扰，如膜电位的改变（电压门控通道）或神经递质的结合（递质门控通道）。

K^+ 选择性漏电通道在决定大多数哺乳动物的跨质膜静息电位中有重要作用。电压门控阳离子通道负责在电兴奋细胞。例如，神经元和骨骼肌细胞中，产生自放大动作电位。递质门控离子通道将化学信号在化学突触处转换成电信号。兴奋性神经递质，如乙酰胆碱和谷氨酸，开放递质门控阳离子通道，从而使突触后膜朝触发动作电位的域值水平去极化。抑制性神经递质，如 GABA 和甘氨酸，开放递质门控 Cl^- 通道或 K^+ 通道，由此通过保持突触后膜的极化状态而抑制触发动作电位。谷氨酸门控离子通道的一个亚类（NMDA 受体通道）对 Ca^{2+} 具有高度渗透性，能引起突触的长期变化，一般认为这种变化与某种形式的学习和记忆有关。

离子通道以复杂的方式进行合作，用以控制电兴奋细胞的行为。例如，一个典型的神经元接收数千种兴奋性和抑制性输入，这些兴奋通过空间叠加和时间叠加结合在一起，在细胞体内产生一种突触后电位（PSP）。PSP 的强度通过轴丘膜中的各种阳离子通道被翻译成动作电位的发放速率。

参考文献

一般文献

Martonosi AN (ed) (1985) The Enzymes of Biological Membranes, vol 3: Membrane Transport, 2nd edn. New York: Plenum Press.

Stein WD (1990) Channels, Carriers and Pumps: An Introduction to Membrane Transport. San Deigo: Academic Press.

Principles of Membrane Transport

Higgins CF (1999) Membrane permeability transporters and channels: from disease to structure and back. *Curr. Opin. Cell Biol.* 11, 495.

Sansom MS (1998) Models and simulations of ion channels and related membrane proteins. *Curr. Opin. Struct. Biol.* 8, 237–244.

Tanford C (1983) Mechanism of free energy coupling in active transport. *Annu. Rev. Biochem.* 52, 379–409.

Carrier Proteins and Active Membrane Transport

Almers W & Stirling C (1984) Distribution of transport proteins over animal cells. *J. Membr. Biol.* 77, 169–186.

Ames GFL (1986) Bacterial periplasmic transport systems: structure, mechanism and evolution. *Annu. Rev. Biochem.* 55, 397–425.

Auer M, Scarborough GA & Kühlbrandt W (1998) Three-dimensional map of the plasma membrane H^+-ATPase in the open conformation. *Nature* 392, 840–843.

Boyer PD (1997) The ATP synthase—a splendid molecular machine. *Annu. Rev. Biochem.* 66, 717–749.

Carafoli E & Brini M (2000) Calcium pumps: structural basis for and mechanism of calcium transmembrane transport. *Curr. Opin. Chem. Biol.* 4, 152–161.

Dean M & Allikmets R (1995) Evolution of ATP-binding cassette trans-

porter genes. *Curr. Opin. Genet. Dev.* 5, 779–785.

Graf J & Haussinger D (1996) Ion transport in hepatocytes: mechanisms and correlations to cell volume, hormone actions and metabolism. *J. Hepatol.* 24(Suppl 1), 53–77.

Henderson PJ (1993) The 12-transmembrane helix transporters. *Curr. Opin. Cell Biol.* 5, 708–721.

Higgins CF (1992) ABC transporters: from microorganisms to man. *Annu. Rev. Cell Biol.* 8, 67–113.

Junge W, Lill H & Engelbrecht S (1997) ATP synthase: an electrochemical transducer with rotatory mechanics. *Trends Biochem. Sci.* 22, 420–423.

Kaback HR, Voss J & Wu J (1997) Helix packing in polytopic membrane proteins: the lactose permease of *Escherichia coli*. *Curr. Opin. Struct. Biol.* 7, 537–542.

Kühlbrandt W, Auer M & Scarborough GA (1998) Structure of the P-type ATPases. *Curr. Opin. Struct. Biol.* 8, 510–516.

Linton KJ & Higgins CF (1998) The *Escherichia coli* ATP-binding cassette (ABC) proteins. *Mol. Microbiol.* 28, 5–13.

Lodish HF (1988) Anion-exchange and glucose transport proteins: structure, function and distribution. *Harvey Lect.* 82, 19–46.

Romero MF & Boron WF (1999) Electrogenic Na^+/HCO_3^- cotransporters: cloning and physiology. *Annu. Rev. Physiol.* 61, 699–723.

Saier MH, Jr (2000) Families of transmembrane sugar transport proteins. *Mol. Microbiol.* 35, 699–710.

Scarborough GA (1999) Structure and function of the P-type ATPases. *Curr. Opin. Cell Biol.* 11, 517–522.

Weber J & Senior AE (2000) ATP synthase: what we know about ATP hydrolysis and what we do not know about ATP synthesis. *Biochim. Biophys. Acta* 1458, 300–309.

Ion Channels and the Electrical Properties of Membranes

Armstrong C (1998) The vision of the pore. *Science* 280, 56–57.

Betz H (1990) Ligand-gated ion channels in the brain: the amino acid receptor family. *Neuron* 5, 383–392.

Biggin PC, Roosild T & Choe S (2000) Potassium channel structure: domain by domain. *Curr. Opin. Struct. Biol.* 10, 456–461.

Choe S, Kreusch A & Pfaffinger PJ (1999) Towards the three-dimensional structure of voltage-gated potassium channels. *Trends Biochem. Sci.* 24, 345–349.

Doyle DA, Cabral JM, Pfuetzner RA et al. (1998) The structure of the potassium channel: molecular basis of K^+ conduction and selectivity. *Science* 280, 69–77.

Franks NP & Lieb WR (1994) Molecular and cellular mechanisms of general anaesthesia. *Nature* 367, 607–614.

Hall ZW (1992) An Introduction to Molecular Neurobiology, pp 33–178. Sunderland, MA: Sinauer.

Hille B (1992) Ionic Channels of Excitable Membranes, 2nd edn. Sunderland, MA: Sinauer.

Hodgkin AL & Huxley AF (1952) A quantitative description of membrane current and its application to conduction and exictation in the nerve. *J. Physiol.* 117, 500–544.

Hodgkin AL & Huxley AF (1952) Currents carried by sodium and potassium ions through the membrane of the giant axon of Loligo. *J. Physiol. (Lond.)* 117, 449–472.

Jessell TM & Kandel ER (1993) Synaptic transmission: a bidirectional and self-modifiable form of cell-cell communication. *Cell* 72(Suppl), 1–30.

Kandel ER, Schwartz JH & Jessell TM (2000) Principles of Neural Science, 4th edn. New York: McGraw-Hill.

Karlin A (1991) Exploration of the nicotinic acetylcholine receptor. *Harvey Lect.* 85, 71–107.

Katz B (1966) Nerve, Muscle and Synapse. New York: McGraw-Hill.

Kim JH & Huganir RL (1999) Organization and regulation of proteins at synapses. *Curr. Opin. Cell Biol.* 11, 248–254.

MacKinnon R, Cohen SL, Kuo A et al. (1998) Structural conservation in prokaryotic and eukaryotic potassium channels. *Science* 280, 106–109.

Malenka RC & Nicoll RA (1999) Long-term potentiation—a decade of progress? *Science* 285, 1870–1874.

Neher E & Sakmann B (1992) The patch clamp technique. *Sci. Am.* 266(3), 28–35.

Nicholls JG, Fuchs PA, Martin AR & Wallace BG (2000) From Neuron to Brain, 4th edn. Sunderland, MA: Sinauer.

Numa S (1989) A molecular view of neurotransmitter receptors and ionic channels. *Harvey Lect.* 83, 121–165.

Perozo E, Cortes DM & Cuello LG (1999) Structural rearrangements underlying K^+-channel activation gating. *Science* 285, 73–78.

Roux B & MacKinnon R (1999) The cavity and pore helices in the KcsA K^+ channel: electrostatic stabilization of monovalent cations. *Science* 285, 100–102.

Sargent PB (1993) The diversity of neronal nicotinic acetylcholine receptors. *Annu. Rev. Neurosci.* 16, 403–443.

Sather WA, Yang J & Tsien RW (1994) Structural basis of ion channel permeation and selectivity. *Curr. Opin. Neurobiol.* 4, 313–323.

Seeberg PH (1993) The molecular biology of mammalian glutamate receptor channels. *Trends Neurosci.* 16, 359–365.

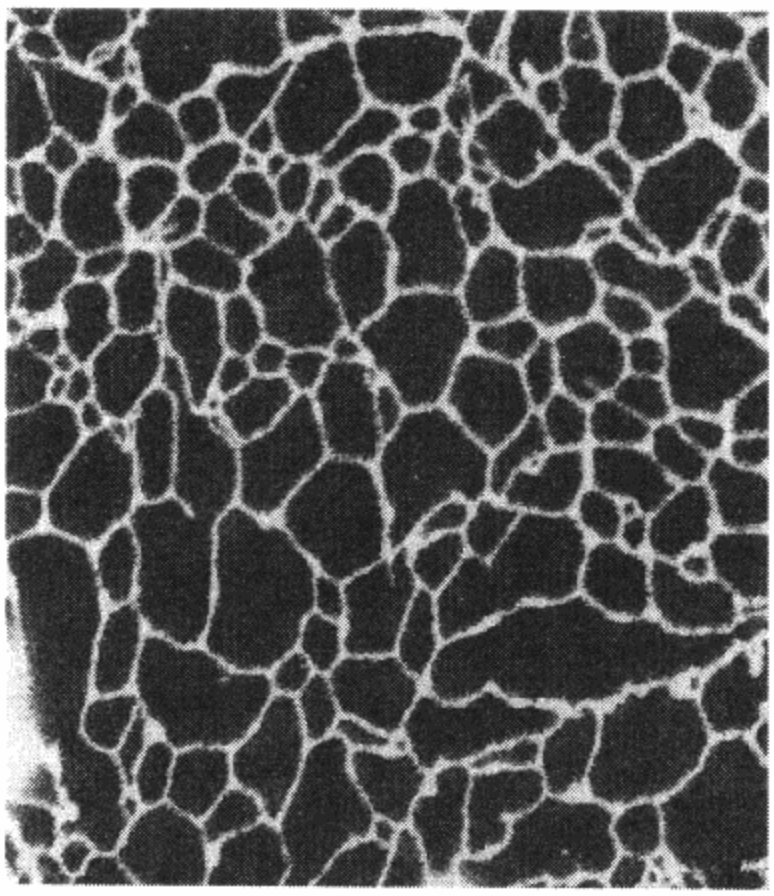
Snyder SH (1996) Drugs and the Brain. New York: WH Freeman/Scientific American Books.

Stevens CF & Sullivan J (1998) Synaptic plasticity. *Curr. Biol.* 8, R151–R153.

Tsien RW (1988) Multiple types of neuronal calcium channels and their selective modulation. *Trends Neurosci.* 11, 431–438.

Unwin N (1998) The nicotinic acetylcholine receptor of the Torpedo electric ray. *J. Struct. Biol.* 121, 181–190.

Utkin Yu N, Tsetlin VI & Hucho F (2000) Structural organization of nicotinic acetylcholine receptors. *Membr. Cell Biol.* 13, 143–164.



内质网（ER）。在这幅烟草细胞中，显示了一个与绿色荧光蛋白标记结合的内质网固有蛋白。它揭示了内质网片层和小管在细胞内的复杂的皮层网络结构。（经 Petra Boevink 和 Chirs Hawes 的允许）

12　细胞内区域和蛋白质分选

细胞内的区域化
细胞核和细胞质间的分子运输
线粒体和叶绿体的蛋白质输入
过氧化物酶体
内质网

细菌可以被看作是被质膜包围的仅有一个胞内分隔的生命体。与其不同，真核细胞胞内被精细地划为多个膜包被的功能特异性空间分隔。每一个胞内空间分隔，或者**细胞器**，都含有一套独特的酶及其他特异性分子。复杂的分配系统可将特异性的产物从一个空间分隔运送到另一个分隔。为了解真核细胞，掌握每一空间分隔中发生了什么，分子如何在分隔间运动，分隔如何产生并维持自身的存在等问题是十分必要的。

蛋白质赋予每个空间分隔特异的结构和特殊的功能。它们催化发生在各个细胞器的化学反应，并且选择性地将小分子运入或运出细胞器**内腔**。蛋白质还是细胞器表面特异性的分子标记，引导新的蛋白质和脂类分子运送到正确的细胞器中。

一个动物细胞包含 100 亿（10^{10}）个蛋白质分子，可分为10 000～20 000种，这些蛋白质分子几乎都是从细胞质基质中开始合成的。每一种新合成的蛋白质都被特异性地运送到需要它的分隔中去。细胞内的蛋白质运输将是本章和下一章的主题。通过追踪蛋白质在胞内空间分隔间运输的通路，可以使我们对错综复杂的细胞内膜系统有一个清晰的了解。

细胞内的区域化

这个介绍性的部分将对胞内分隔和它们彼此之间的关系作一个简要的概述。为此，我们在概念上将细胞器划分成少数几个独立的家族，并且讨论蛋白质是如何被引导至特定细胞器，如何穿过细胞器膜的。

所有真核细胞具有一套相同的膜包被基本细胞器

很多极其重要的生物化学过程都发生在膜表面或膜内。例如，脂类的代谢过程就主要是由结合于膜上的酶催化进行的。此外，氧化磷酸化和光合磷酸化也需要膜偶联 H^+ 的运输和 ATP 的合成过程。细胞内膜系统不仅增加了细胞的膜表面积，还构造了和细胞质基质分离的封闭胞内分隔，从而为细胞提供了功能特异性的溶液环境。由于细胞器膜的类脂双层膜对于大多数亲水性的分子是不通透的，因此各细胞器膜上必须有膜转运蛋白，这些蛋白质负责将特定的代谢物运入或运出细胞器。此外，细胞器膜还应具有能转入该细胞器特异性蛋白质的机制。

图 12-1 显示了所有真核细胞内都具有的主要胞内分隔。细胞核包含细胞的主要基因组，是 DNA 和 RNA 合成的主要场所。细胞核周围的**细胞质**由细胞质基质和悬浮于其中的胞质细胞器组成。**细胞质基质**约占细胞总体积的一半以上，是蛋白质合成和降解的场所，同时也是细胞中间代谢反应的场所，很多构建生物大分子的小分子物质是在细胞质基质中降解或合成的（讨论见第 2 章）。

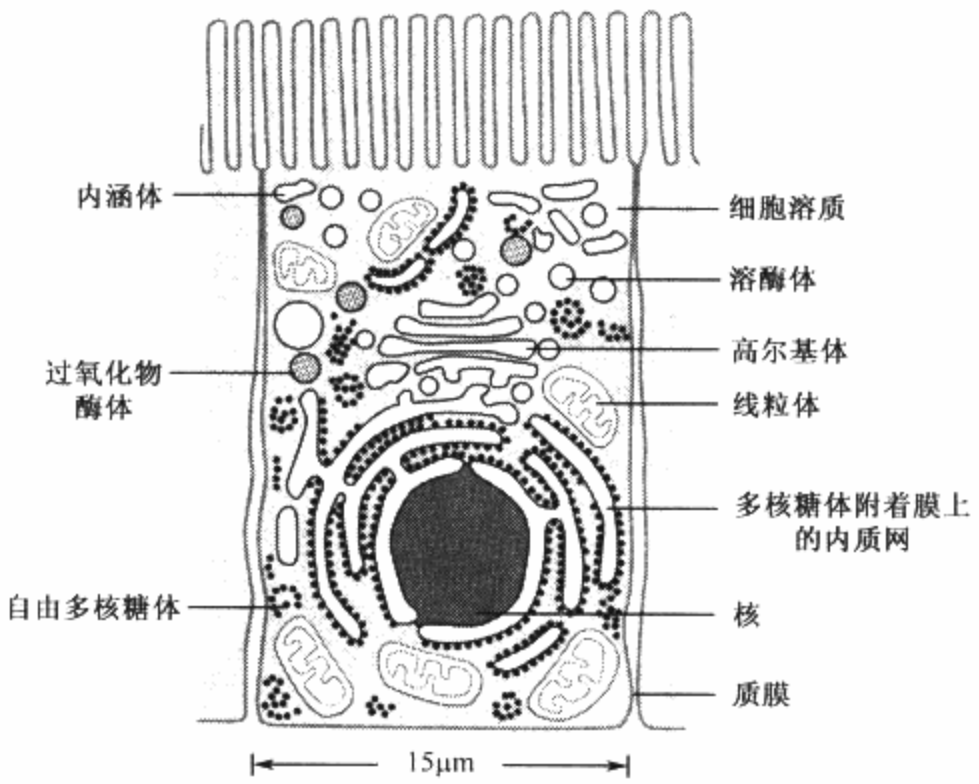


图 12-1　动物细胞的主要胞内分隔

细胞质基质（灰色）、内质网、高尔基体、细胞核、线粒体、胞内体、溶酶体和过氧化物酶体是不同的胞内分隔。它们至少由一层选择性透过的膜包围，使之与细胞的其他部分分开

大约一半真核细胞内膜系统包围形成了错综复杂的内质网结构（ER）。一些内质网膜的胞质面上结合着很多核糖体，它们负责可溶蛋白质和膜内在蛋白质的合成，这些蛋白质中的很大一部分将会被分泌到细胞外或者其他细胞器。在下文，我们将会看到输入内质网的蛋白质是边合成边进行转运的，而输入其他细胞器的蛋白质是在合成过程完成后才开始转运的。这解释了内质网为何是唯一结合有核糖体的细胞器。此外，内质网为细胞生产大部分的脂类，同时还是细胞的 Ca^{2+} 库。内质网将其合成的大部分蛋白质和脂类运送至高尔基体。高尔基体是由一组被称作高尔基池的盘状膜囊有序罗列形成的细胞器，它接收来自内质网的脂类和蛋白质，再将它们发往不同的目的地；并在这个过程中不断对脂类和蛋白质分子进行共价修饰。

线粒体和叶绿体（仅在植物中）产生大部分 ATP，这些 ATP 为细胞驱动各种反应提供自由能；叶绿体是质体中较为特殊的一种，在植物细胞中还承担着诸如营养物质和色素的储存等其他功能。溶酶体含有多种消化酶，可以降解细胞内失效或死亡的细胞器和细胞经胞吞作用摄入体内的大分子及囊泡。但在进入溶酶体之前，胞吞物质必须首先经过一系列称作胞内体的细胞器。过氧化物酶体是一种包含多种氧化酶的囊泡状细胞器。

大体上说，各种类型细胞中的同一种细胞器行使的基本功能相同，但为了使特定的细胞获得特殊的功能，不同细胞中相同细胞器的数量及其附加的功能都是互不相同的。

膜封闭的胞内分隔的总和通常几乎占细胞体积的一半（表 12-1），因而需要大量的细胞内膜结构。以肝细胞和胰腺细胞为例，它们的内质网膜面积分别约是其质膜体积的25倍和12

表 12-1　肝细胞中各主要胞内分隔的相对体积

细胞器	占细胞总体积百分比/%
细胞质基质	54
线粒体	22
粗面内质网池	9
光面内质网池及高尔基体	6
细胞核	6
过氧化物酶体	1
溶酶体	1
胞内体	1

表 12-2　两种真核细胞中各种膜的相对数量

膜的类型	占细胞总膜量的百分比/%	
	肝细胞	胰腺细胞
细胞膜	2	5
粗面内质网膜	35	60
光面内质网膜	16	<1
高尔基体膜	7	10
线粒体		
外膜	7	4
内膜	32	17
细胞核		
内膜	0.2	0.7
分泌囊泡膜	待定	3
溶酶体膜	0.4	待定
过氧化物酶体膜	0.4	待定
胞内体膜	0.4	待定

注：这两种细胞的大小有很大差别：肝细胞的平均体积约是 5000μm^3，而胰腺细胞的平均体积是 1000μm^3。细胞总膜面积大约分别是110 000μm^2和13 000μm^2。

倍（表 12-2）。从面积和质量上看，大部分真核细胞的质膜仅仅是次要的膜结构（图 12-2）。

包被膜的细胞器通常在细胞质基质中有固定的位置。例如，在大多数细胞中，高尔基体通常很靠近细胞核，而内质网的管腔则是从细胞核附近出发，延伸至整个细胞质基质。细胞器的特异性定位是由细胞器和细胞骨架间的相互作用决定的。例如，内质网和高尔基体的定位就需要依赖完整无损的微管列阵，如果微管被某种试剂解聚，那么高尔基体将会断裂成碎片并且弥散在整个细胞中，而内质网将会朝向细胞中心崩解（讨论见第 16 章）。

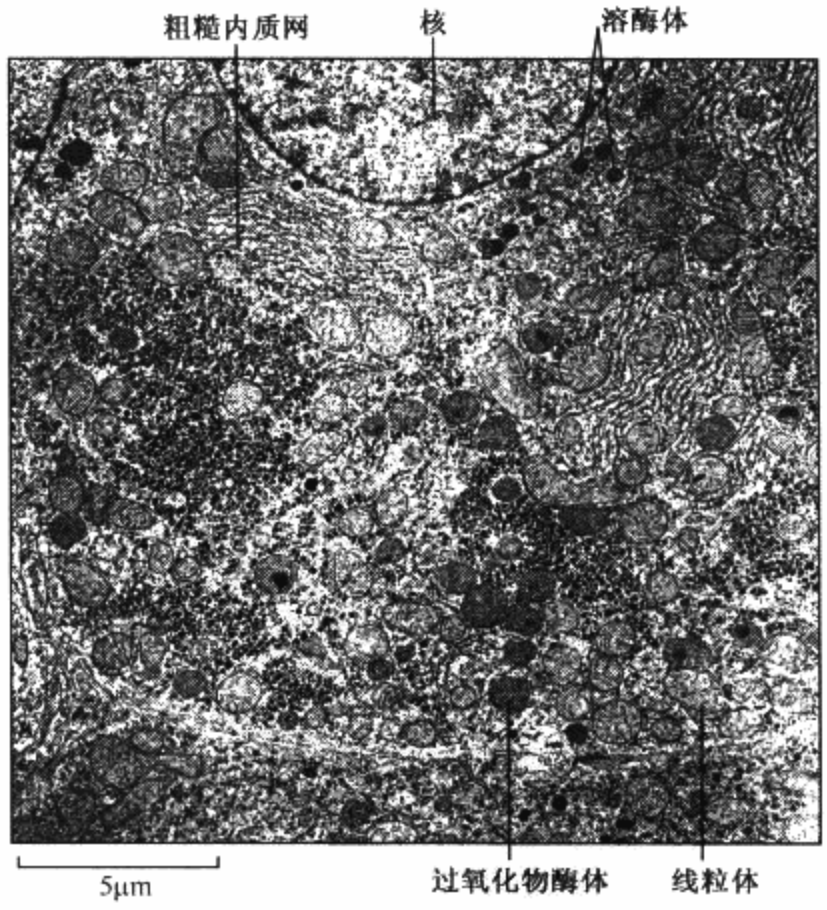


图 12-2　肝细胞横断面电镜照片局部

显示大部分主要的胞内分隔结构

膜包被的细胞器的形态学关系可以从进化发生的角度得到解释

了解胞内分隔（即细胞器）可能的进化途径对于理解它们之间的关系是非常有帮助的。最早的真核细胞的前身被认为是和细菌相似的简单生物体，拥有细胞膜却没有内膜系统，因此这种细胞的细胞膜必须行使所有依赖膜的功能，包括泵出或泵入离子、ATP 的合成、蛋白质的分泌和脂类的合成。现存的典型真核细胞和典型细菌（如 *E. coli*）相比，前者的长度是后者的 10～30 倍，体积是后者的 1000～10 000倍，因而真核细胞的比表面积比细菌小得多。由于真核细胞的细胞膜太小以至于不足以承担如此多的膜功能，因而广泛分布的内膜系统就十分必要。因此，大量内膜的存在也可以看作是对细胞体积增加的一种适应。

显然，内膜的进化伴随着膜功能的特化。以叶绿体中类囊体的产生为例，这些类囊体是在植物叶片中的前质体转变为叶绿体的过程中产生的。前质体是一种小型的细胞器前体，存在于所有未成熟的植物细胞中。它由双层膜包被，并且根据细胞分化的需要转变分化。例如，叶肉细胞中的前质体转化为叶绿体，而在其他类型的细胞中前质体则转化为储存淀粉、脂肪和色素的细胞器［图 12-3(A)］。在分化为叶绿体的过程中，前质体内层膜上会出现一些特化的膜结构并从内层膜上分离下来，形成一种特化的囊泡结构，即类囊体。类囊体含有叶绿体进行光合作用所需的整套结构［图 12-3(B)］。

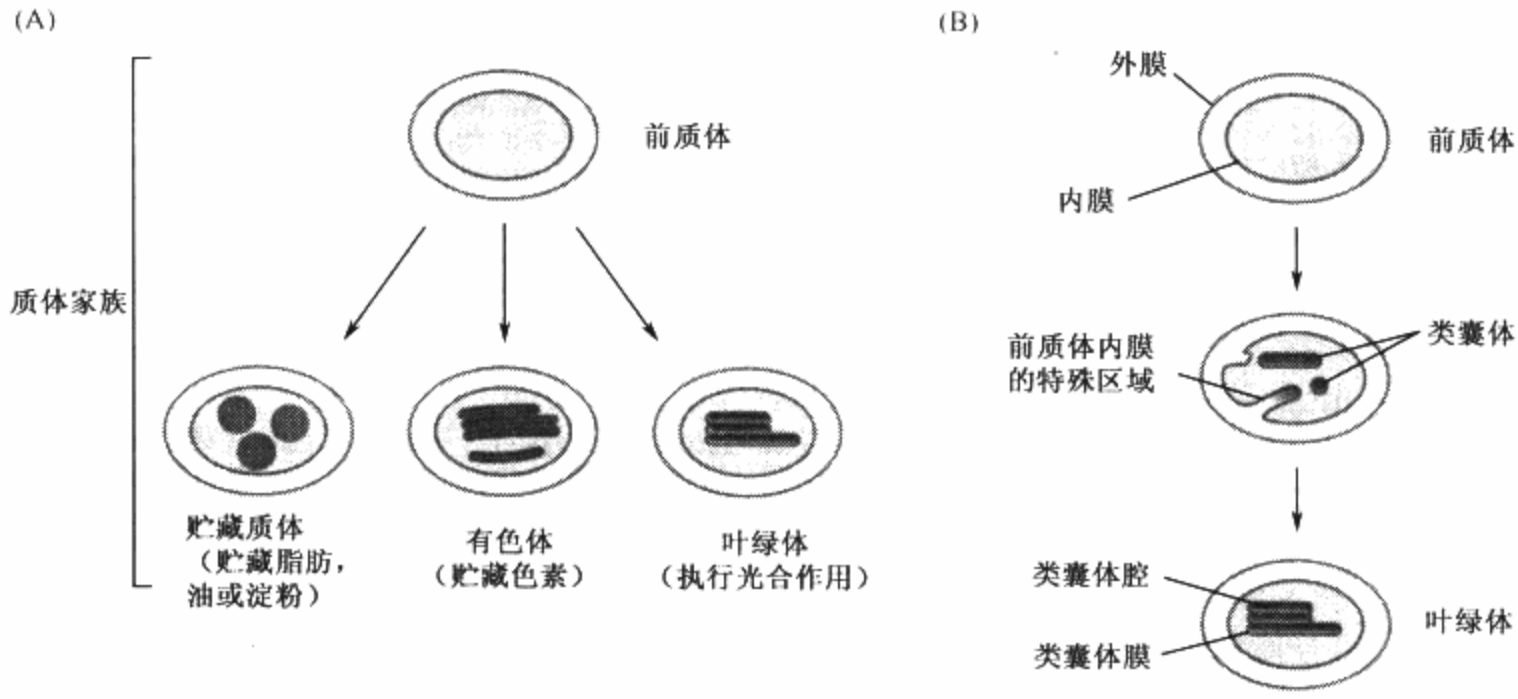


图 12-3　质体的转化

（A）前质体和植物卵细胞的细胞质共遗传。当未成熟的植物细胞分化时，前质体根据特定细胞的需要转化为叶绿体（在绿色叶肉细胞中）、积累淀粉（如土豆块茎中）或油和脂滴（如脂肪种子中）的储藏质体，或储存色素的有色体（如花瓣中）。(B) 类囊体的发生。在叶绿体形成过程中，前质体内膜上的特化膜片段内折脱落形成类囊体囊泡，进而形成成熟的类囊体。类囊体膜包被分隔一个独立的空间，类囊体腔。它在结构和功能上都和叶绿体的其他部分不同。在叶绿体繁殖的过程中，类囊体可以自发的生长分裂

真核细胞中的其他胞内分隔可能也是通过相似的方式起源发生的［图 12-4(A)］。从细胞膜上分离的特化内膜结构，会使细胞器的内外具有相同的形态结构。事实上，包括内质网、高尔基体、胞内体和溶酶体在内的所有和分泌、内吞途径相关的细胞器都具有这种形态学上的相似关系，因此我们推断这些细胞器可能是同一家族的成员。下一章将会详细地讨论这些细胞器如何进行广泛的物质信息交流，以及它们如何通过不断出芽、融合的转运泡和细胞外环境联系（图 12-5）。

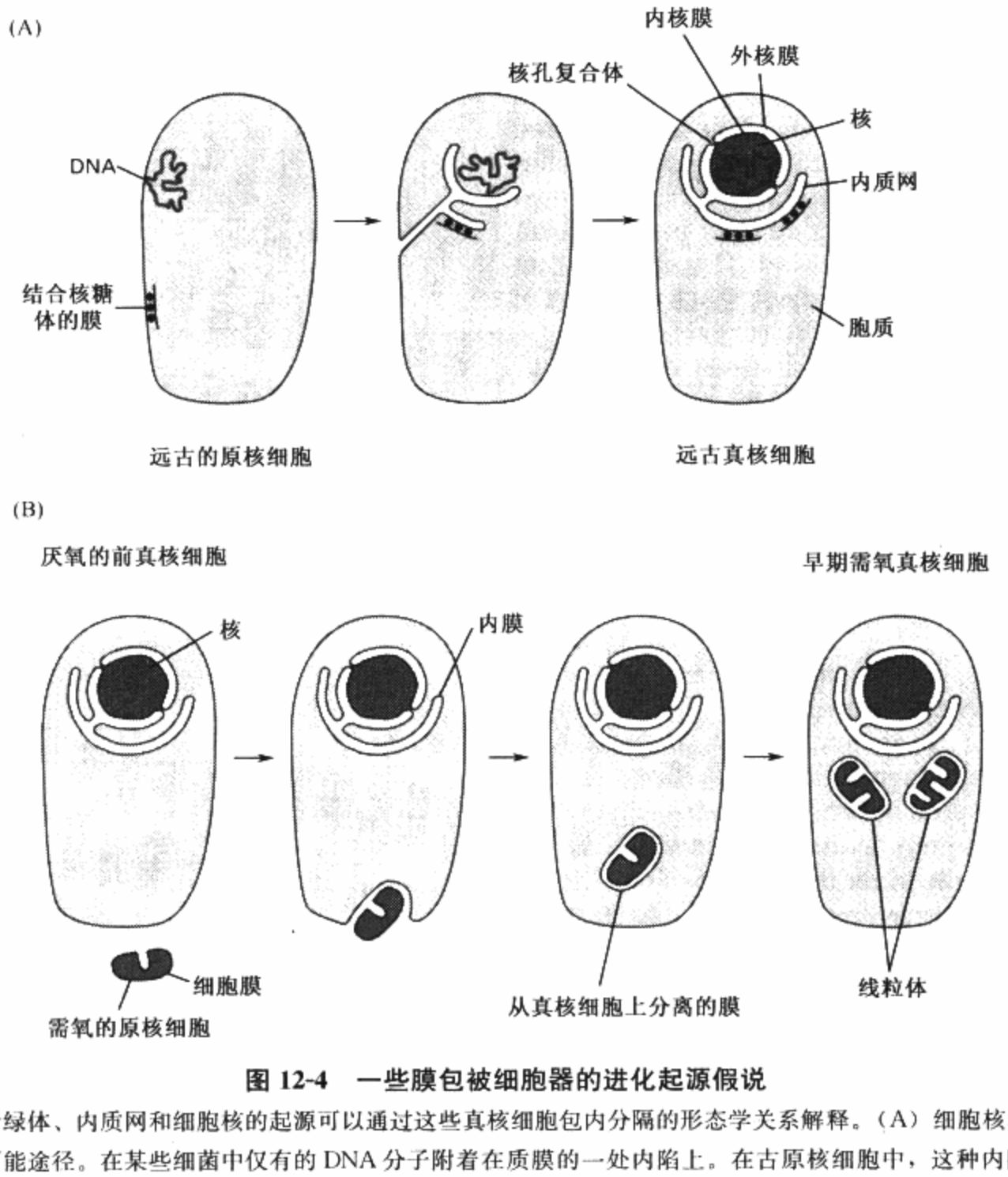


图 12-4　一些膜包被细胞器的进化起源假说

线粒体、叶绿体、内质网和细胞核的起源可以通过这些真核细胞包内分隔的形态学关系解释。（A）细胞核和内质网进化的一条可能途径。在某些细菌中仅有的 DNA 分子附着在质膜的一处内陷上。在古原核细胞中，这种内陷可以重排形成围绕 DNA 的被膜，同时仍然允许 DNA 接触细胞质基质（因为 DNA 必须指导蛋白质的合成）。这层膜最终完全从质膜上分离下来，形成一个由双层膜包被的细胞核。如图所示，核被膜核孔复合体穿通。由于细胞核被两层在核孔处连续的膜包围，细胞核在形态结构上和细胞质基质一致。事实上，在有丝分裂过程中细胞核物质和细胞质混合。内质网腔与核的内外膜间隙相通，并且和细胞外间隙的形态结构一致。（B）线粒体（和质体）被认为起源于被大型前真核细胞吞噬的细菌。它们保留着自身的自主性。这可以解释为什么两种细胞器的内腔独立于连接许多其他细胞器内腔的膜运输系统之外

在第 14 章，我们将讨论线粒体、质体因含有自己的基因组而和其他有膜细胞器有所不同。这些基因组的特征以及两种细胞器蛋白质和一些现存细菌蛋白质极其相似的事实，强有力地说明线粒体和质体很可能是从某些细菌进化而来的。这些细菌起初和一些细胞共栖，进而被后者吞食成为其结构的一部分。根据图 12-4(B) 所示的假说，线粒体和质体的内层膜即原始的细菌的细胞膜，而其内腔则从细菌的细胞质基质进化而来。因为两种细胞器都是内吞起源，因而它们被双层膜包围，独立于连接大部分有膜细胞器和细胞外界的膜泡运输系统之外。

以上讨论的进化假说将真核细胞的胞内分隔分成了 4 个家族。①细胞核和细胞质基质：它们通过核孔复合体相连，在形态结构上是连续的（在功能上是相互独立的）；②在分泌和内吞途径中发挥作用的所有细胞器：包括内质网、高尔基体、胞内体、溶酶体、各类转运中间物（如转运泡）以及过氧化物酶体；③线粒体；④质体（仅存在于植物中）。

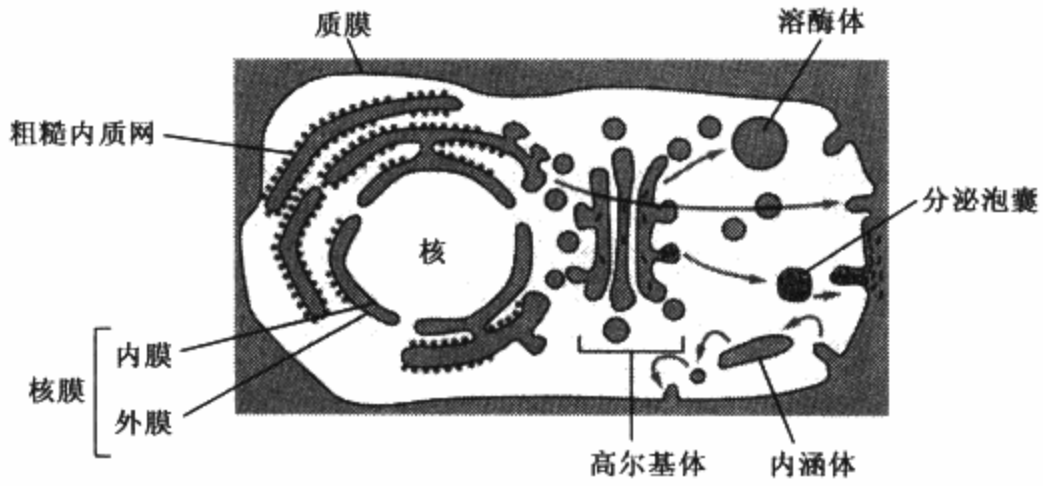


图 12-5　真核细胞分泌和胞吞途径中所涉及细胞器的形态结构关系

形态结构相同的空间用红色表示。大体而言，膜的出芽和融合周期使得所有这些细胞器之间以及细胞器和细胞外环境之间可以通过运输囊泡相联系。蓝色箭头指示了运输网络中的输出和输入途径，这将在第 13 章讨论。某些细胞器，如线粒体和质体（仅在植物细胞中）不参加图中所示的运输系统

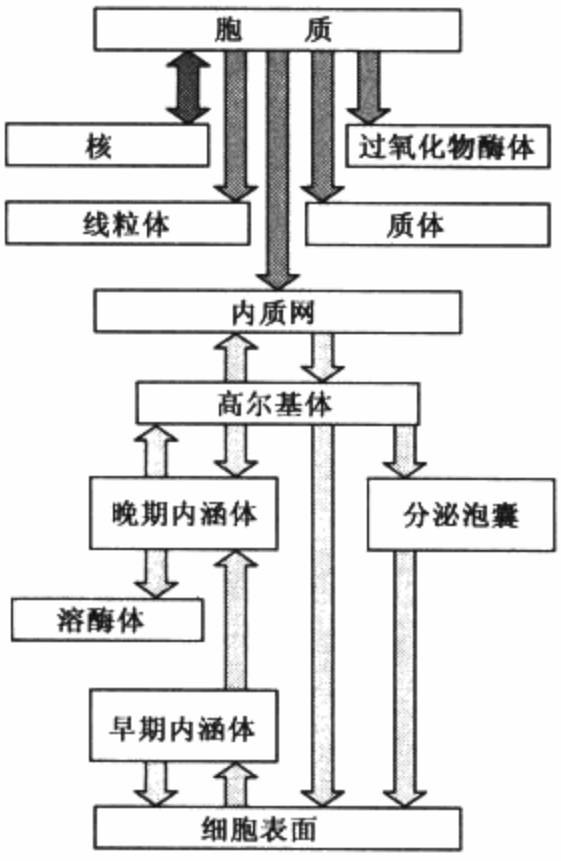


图 12-6　蛋白质运输的简单路线图

蛋白质可以通过门控运输（红色）、跨膜运输（蓝色）或囊泡运输（绿色）从一个胞内分隔进入另一个分隔。特定蛋白质的氨基酸序列中含有引导该蛋白质通过此系统某一路径进而决定蛋白质在细胞中最终位置的信号序列。整个过程从蛋白质在细胞质基质核糖体中的合成开始，到蛋白质达到最终目的地结束。每一个中间环节（框）都会决定蛋白质是被保留在这个细胞器中还是继续转运。原则上说，无论是驻留还是输出都需要一个信号。我们会把这张图作为向导在本章和下一章中反复使用。颜色标记的路径将被重点讨论

蛋白质可以在各分隔之间以不同的方式运动

除了少数蛋白质在线粒体和质体内的核糖体上合成，几乎所有的蛋白质都是在细胞质基质中的游离核糖体上开始合成的。这些蛋白质的命运将由它们自身的氨基酸序列决定，序列中的**分选信号**将引导蛋白质到细胞质基质外的其他位置。大部分蛋白质没有分选信号，因此它们只能永久停留在细胞质基质中。然而那些含有特定分选信号的蛋白质将从细胞质基质被运送到细胞核、内质网、线粒体、质体或过氧化物酶体。分选信号还可以引导蛋白质经内质网被运送到细胞内的其他目的地。

为了理解分选信号是如何发挥功能的，必须区分蛋白质在胞内分隔间运动的三种基本途径。图 12-6 概述了这三条途径的各个步骤，三种机制具体描述如下，前两种途径将在本章详细阐述，第三种途径将在第 13 章中讨论。

1. **门控转运**：蛋白质在细胞质基质和细胞核之间的转运实际上是经核孔复合体在一个形态结构一致的连续空间内转运的过程。核孔复合体好像一个具有选择性的闸门，主动转运特定的大分子及其相似物，同时也允许小分子自由扩散。

2. **跨膜转运**：膜结合的转运蛋白可以直接将特定的蛋白质从细胞质基质跨膜运送至另一个

形态结构完全不同的空间中。被转运的蛋白质分子一般是以伸展的线性状态从转运蛋白中通过。细胞质基质蛋白被运送至内质网管腔和线粒体的最初过程就是经由这种途径实现的。

3. **膜泡转运**：膜包被的转运中间物（可能是小球状的转运泡，或较大的、不规则的细胞器残片）可以把蛋白质从一个分隔运送至另一个分隔。这些转运泡或残片从原先的细胞器膜上分离时就从该细胞器腔内携带上待转运蛋白质，当转运中间物和另一个细胞器融合时，便将蛋白质卸载到这个细胞器内（图 12-7）。可溶性蛋白质从内质网到高尔基体的转运便是通过这种途径。因为被转运的蛋白质不跨膜，因而膜泡转运只能让蛋白质在形态结构相同的胞内分隔间运动(图 12-5)。我们将在第 13 章进一步讨论膜泡运输。

这 3 种转运方式通常都由被转运蛋白中的分选信号引导发生。在转运过程中，分选信号会被互补的受体蛋白识别。例如，一个大蛋白质要转运进入细胞核，那么这个蛋白质必须含有可被受体蛋白识别的分选信号，这样受体蛋白才能引导它穿过核孔复合体。如果一种蛋白质要进行跨膜转运，那么它必须含有能让膜结合转运蛋白识别的分选信号。同理，如果一种蛋白质要被装载到特定的膜泡或细胞器中，它的分选信号必须能被特定膜上的互补受体识别。

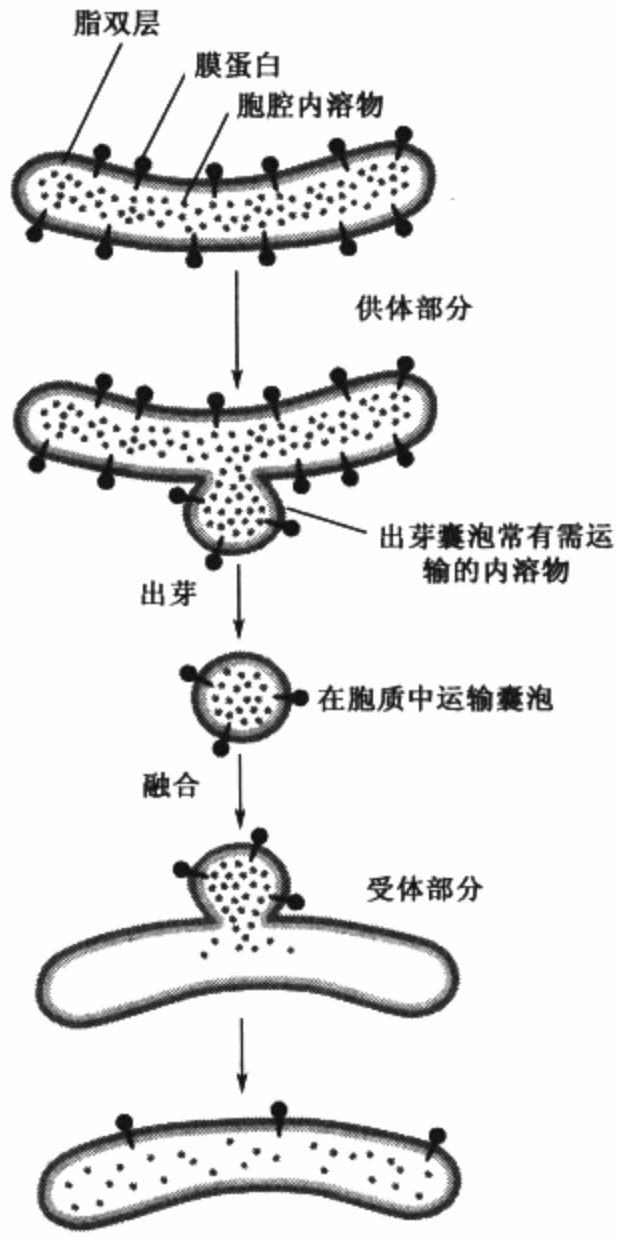


图 12-7　囊泡运输过程中的出芽和融合

运输囊泡从一个细胞器（供体）出芽和另一个细胞器（目标受体）融合。在这个过程中，可溶成分（红色点）从一个内腔被转移到另一个内腔。注意膜也同时被转运，共体膜的蛋白质和脂类方向仍在目标受体膜中保持。因此，膜蛋白始终维持不对称定向，朝向细胞质基质的总是同一个结构域

信号序列和信号斑引导蛋白质至细胞内的正确位置

至少存在两种蛋白质分选信号。一种在于舒展的连续氨基酸序列中，一般具有 15～60 个残基。一旦分选完成，蛋白质被运送至目的地，一些**信号序列**将被特异性的**信号肽酶**从成熟蛋白上切除。另一种分选信号是在蛋白质折叠完成后，其表面的原子按一定的三维结构排布形成的。形成这种**信号斑**的氨基酸残基可能在线性氨基酸序列上相距很远，并且它们会被保留在成熟蛋白中，不会被切除（图 12-8）。信号序列用于将蛋白质从细胞质中引导至内质网中、线粒体、叶绿体、过氧化物酶体中，也用于将蛋白质从细胞核中引导至细胞质中、从高尔基体引导至内质网中。引导蛋白质从细胞质进入细胞核的分选信号可以是短信号序列或者是较长的易折叠成信号斑的序列。此外，信号斑还可以引导新合成的降解酶进入溶酶体中。

每一种信号序列都具体对应细胞内的一个特定目的地。最初被运送至内质网的蛋白质通常在 N 端具有由 5～10 个疏水氨基酸组成的信号序列。这些蛋白质继而从内质网被运送至高尔基体中；但如果在这些蛋白质的 C 端有一个 4 个氨基酸残基组成的特异性序列，它们将被识别为内质网蛋白质而被重新运回内质网中。定位于线粒体的蛋白质具有另一种信号序列，

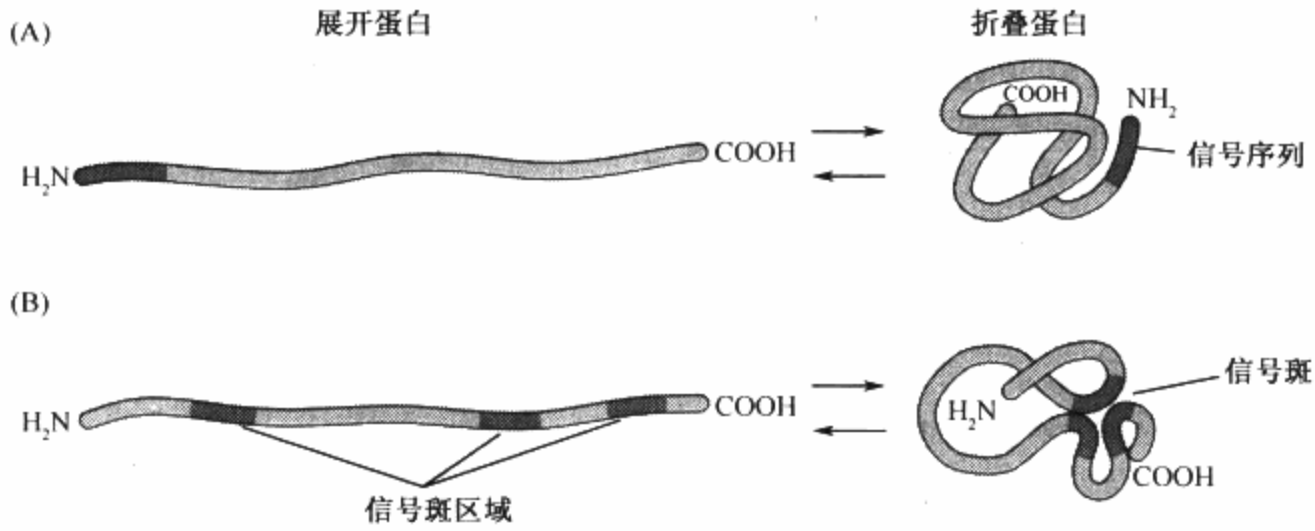


图 12-8　蛋白质中两种分选信号的形成

（A）信号仅存在于氨基酸序列的单一片段上，暴露于折叠蛋白质表面，称为信号序列。信号序列经常处于多肽链的末端（如图所示），但也可处于内部。（B）蛋白质折叠前，处于多肽链上彼此分离的不同位置的氨基酸聚集并置形成信号斑（如图所示）。此外，折叠蛋白表面具有固定距离的分离片段也能形成信号

在这种信号序列中带正电的氨基酸和疏水氨基酸交替排列。定位于过氧化物酶体的蛋白质在其 C 端具有 3 个特征氨基酸组成的信号肽。

表 12-3 列举了一些特异性的信号序列。已有实验支持，这些信号序列在蛋白质定位的过程中发挥了重要作用。实验中利用基因工程技术将信号肽从一种蛋白质转移至另一种蛋白质上。例如，将 N 端内质网信号序列转移至细胞质蛋白的 N 端，则本应在细胞质中的蛋白质转移至内质网中。因此，信号序列的存在是蛋白质的定位的充分必要条件。即使氨基酸顺序有很大的不同，目的地相同的所有蛋白质的信号序列在功能上是等价的，它们的物理性质，如疏水性，在信号识别的过程中往往要比氨基酸序列重要。

表 12-3　一些典型的信号序列

信号序列的功能	信号序列举例
Import into nucleus	-Pro-Pro-Lys-Lys-Lys-Arg-Lys-Val-
Export from nucleus	-Leu-Ala-Leu-Lys-Leu-Ala-Gly-Leu-Asp-Ile-
Import into mitochondria	$^{+}H_3N$-Met-Leu-Ser-Leu-Arg-Gln-Ser-Ile-Arg-Phe-Phe-Lys-Pro-Ala-Thr-Arg-Thr-Leu-Cys-Ser-Ser-Arg-Tyr-Leu-Leu-
Import into plastid	$^{+}H_3N$-Met-Val-Ala-Met-Ala-Met-Ala-Ser-Leu-Gln-Ser-Ser-Met-Ser-Ser-Leu-Ser-Leu-Ser-Ser-Asn-Ser-Phe-Leu-Gly-Gln-Pro-Leu-Ser-Pro-Ile-Thr-Leu-Ser-Pro-Phe-Leu-Gln-Gly-
Import into peroxisomes	-Ser-Lys-Leu-COO^{-}
Import into ER	$^{+}H_3N$-Met-Met-Ser-Phe-Val-Ser-Leu-Leu-Leu-Val-Gly-Ile-Leu-Phe-Trp-Ala-Thr-Glu-Ala-Glu-Gln-Leu-Thr-Lys-Cys-Glu-Val-Phe-Gln-
Return to ER	-Lys-Asp-Glu-Leu-COO^{-}

注：不同信号序列的某些典型特征被颜色标记。对于信号序列的功能特别重要的残基，用红色标记了带正电荷的，绿色标记了带负电荷的。相似地，重要的疏水性氨基酸用黄色标记，重要的羟基化氨基酸用蓝色标记。$^{+}H_3N$ 表示蛋白质的 N 端，COO^{-} 表示 C 端。

信号斑比信号序列更难以分析，所以对它们的结构了解得更少。由于它们的功能往往来源于复杂的蛋白质三维折叠模式，因此很难在维持这种结构的情况下从一种蛋白质转移至另一种蛋白质。

两种分选信号都被互补的分选受体识别，后者引导蛋白质到达目的地后将其卸载于此。受体行使着类似催化剂的功能：当完成一次定位后，它们会回到原来的位置以供再次使用。

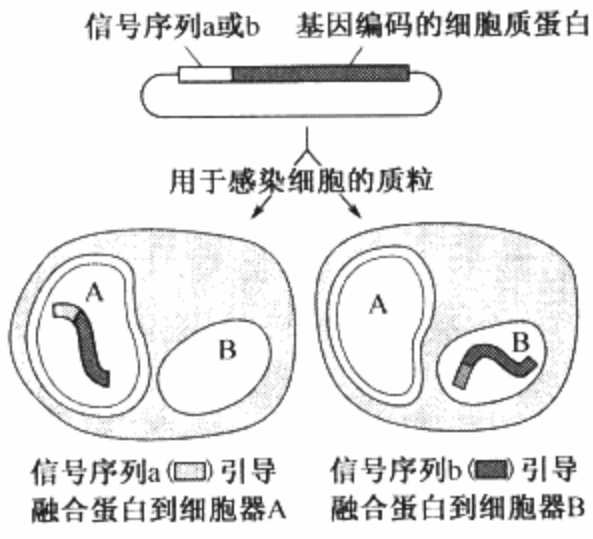
大部分分选受体识别多种蛋白质，因此它们可被看作公共的运输系统，负责将各种成分运送到细胞中的正确位置。

专题 12-1 显示了研究蛋白质如何从细胞质基质中被引导至特定的分隔中以及蛋白质如何跨膜的主要方法。

专题 12-1　研究信号序列和蛋白质跨膜运输的方法

转染方法以确定信号序列

一种显示信号序列是将蛋白质定位于指定包内分隔的充分必要条件的方法是制作融合蛋白：用基因工程技术将信号序列和一个本应驻留在细胞质基质中的蛋白质相连，当编码这个融合蛋白的 cDNA 被转染进入细胞后，用免疫染色或分级分离的方法确定融合蛋白的位置。通过定点诱变改变信号序列，就可以确定序列中哪一部分对信号功能最重要。



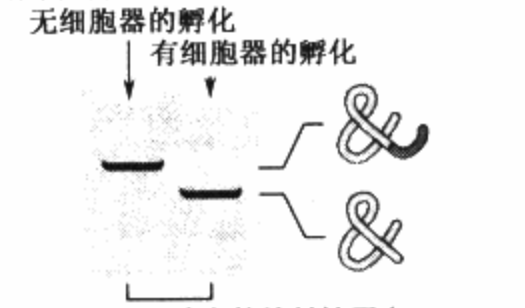
研究蛋白质易位机制的一种生化方法

在这种方法中，一个含有特异性信号序列的标记蛋白质被输入体外已被分离的细胞器。被标记蛋白质通常经由编码蛋白质的分离纯化 mRNA 的体外翻译得到。放射性氨基酸用于标记新合成的蛋白质，以便和已经存在于体外翻译系统中的众多其他蛋白质相分离。

有三种方法经常用于检测被标记的蛋白质是否已经易位进入细胞器：

1. 离心过程中，被标记蛋白质和细胞器共分级分离出来。

2. 信号序列被细胞器内的特异性蛋白酶切除。

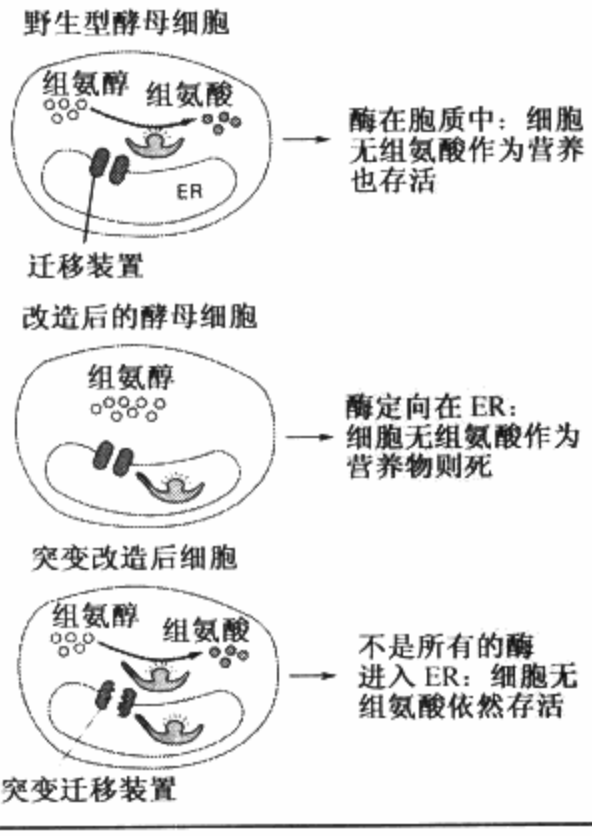


3. 将蛋白酶加入培养基中，蛋白质可以不被降解，但若先加入去污剂破坏细胞器的膜，再加入蛋白酶后蛋白质被降解。

利用这种体外分析方法，可以确定哪些成分（蛋白质、ATP、GTP 等）对于易位是必需的。

研究蛋白质易位机制的遗传学方法

编码易位装置结构成分的基因突变酵母细胞一直是研究蛋白质易位的得力工具。因为不能进行蛋白质跨膜易位的细胞将会死亡，所以实验的关键是设计一种方法，使得仅引起蛋白质易位部分缺陷的弱突变株被分离出来。



一种方法利用基因工程来设计一种特殊的酵母细胞。例如，组胺醇脱氢酶在正常情况下存在于细胞质基质中，催化组胺醇产生必需氨基酸组氨酸。在构建的酵母细胞中，原先的组胺醇脱氢酶基因被一个经过基因工程修饰，添加了信号序列的融合蛋白基因所取代，因而这个酶将被错误运送至内质网（ER）中。当这种细胞在没有组氨酸的环境中生长时，由于所有的组胺醇脱氢酶都被隔离在内质网中而无法被使用，细胞将死亡。然而部分抑制从细胞质基质到内质网蛋白质易位的突变细胞则可以存活下来，因为有足够的组胺醇脱氢酶驻留在细胞质基质中催化组氨酸的产生。通常可以筛选突变蛋白在常温仍具有部分功能而在高温下完全失去功能的细胞。不管有无组氨酸存在，携带有这种温度敏感型突变基因的细胞在高温下都会死亡，因为任何蛋白质都无法被转运进入内质网。这样，通过将携带有酵母基因组 DNA 随机克隆片段的酵母质粒载体转染突变细胞，就可以确定突变基因所对应的正常基因：因为能够使突变细胞在高温下生长的起补救作用的 DNA 片段应编码突变基因对应的野生型蛋白质。

大部分膜包被细胞器并不是从头形成：细胞器自身提供必需信息

当细胞通过分裂增殖时，它必须复制膜包被细胞器。大体而言，细胞通过不断融入新分子来增大细胞器；然后增大的细胞器分裂并被分配到两个子细胞中去。因此，每个子细胞从母细胞中继承了一整套特殊的细胞膜系统。这种继承非常必要，因为细胞不能从头开始形成这些膜系统。例如，如果将内质网从细胞中完全除去，细胞就不能重建它。正如在下文将讨论的，定义内质网并且行使很多功能的膜蛋白是由内质网自身产生的，如果没有先前存在的内质网，或者没有含有选择性输入蛋白质进入内质网的易位体的特定膜（包括内质网特异性易位体自身），就不能形成新的内质网。线粒体、质体和过氧化物酶体也是如此。

因此，构建膜包被细胞器的信息似乎并不仅仅存在于编码细胞器特定蛋白质的 DNA 中。前已存在的细胞器膜蛋白中的表位遗传信息也必须参与到这一构建过程中来，并通过细胞器从亲代细胞传递到子代细胞。这种信息对细胞器的繁殖可能是必需的，如同 DNA 中的信息对细胞核苷酸和氨基酸序列的繁殖的重要性。

正如第 13 章即将详细讨论的，从内质网不断脱落下来的囊泡仅包含特定的内质网蛋白质，因而它们具有和内质网不同的组成成分。相似地，细胞膜也不断产生特殊的胞吞囊泡。因此，某些膜包被细胞器可以从其他细胞器形成，并不一定是在细胞分裂过程中从亲代细胞器遗传继承的。

小结

真核细胞的内膜系统包被了将近一半的细胞体积，将从称为细胞器的胞内分隔独立出来。所有细胞器中都具有的主要膜包被细胞器有内质网、高尔基体、细胞核、线粒体、溶酶体、胞内体和过氧化物酶体；植物细胞还含有质体，如叶绿体。每种细胞器都含有一套特殊的蛋白质来行使自身的特殊功能。

每种新合成的蛋白质必须从合成它们的核糖体进入细胞器行使功能。这一过程通过蛋白质氨基酸序列中的信号序列或信号斑介导的特殊途径完成。信号序列和信号斑互补的分选受体识别，后者将蛋白质定位在指定细胞器中。在细胞质基质中行使功能的蛋白质不含分选信号，因此，合成后始终停留在细胞质基质中。

在细胞分裂过程中，内质网和线粒体等细胞器被完整分配到子细胞中。这些细胞器中所含的某些信息对于它们的构建是必需的，因此细胞器不是从头开始合成的。

细胞核和细胞质间的分子运输

核被膜负责包围 DNA 并限定细胞核的区域，它由两层膜组成，其上穿插有核孔复合体（图 12-9）。虽然内核膜和外核膜是连续的，但其蛋白质成分不同。**内核膜**上有特殊的蛋白质为染色质与核纤层蛋白网络结构的结合提供结合位点，核纤层的作用是为内膜提供结构支持。内核膜外的**外核膜**和内质网形成连续的结构。与这一章的末尾将要讨论的内质网膜一样，外核膜上也布满负责蛋白质合成的核糖体，由这些核糖体合成的蛋白质被运至内核膜和外核膜间的间隙（核周间隙）、核周间隙和内质网腔是连通的（图 12-9）。

细胞质基质和细胞核之间存在持续的双向运输。在细胞核中行使功能的蛋白质，如组蛋白、DNA 和 RNA 聚合酶、基因调节蛋白和 RNA 转录后加工蛋白——在细胞质基质中产生，被选择输入细胞核。同时，tRNA 和 mRNA 在细胞核中合成，然后被输出进入细胞质基质。

和输入过程一样，输出过程也是选择性的；例如，mRNA 只有经核内 RNA 转录后加工的正确修饰才能被输出。一些情况下运输过程十分复杂。例如，在细胞质基质中合成的核糖体蛋白质被输入细胞核，然后和细胞核内新合成的核糖体 RNA 装配形成颗粒，再作为核糖体亚基的一部分输入细胞质基质。这其中的每一步都需要核被膜的选择性运输。

核孔复合体贯穿核被膜

所有真核细胞的核被膜上都布满一种称为**核孔复合体**的大型精巧结构，并被其贯穿。动物细胞每个核孔复合体的相对分子质量是125 000 000，由至少50 种不同的蛋白质组成。这些蛋白质称为**核孔蛋白**，在核孔处以八面体对称形式排布（图 12-10）。

一般来说，细胞核内的转录越活跃，核被膜上的核孔复合体数就越多。一个典型的哺乳动物细胞核被膜上有 3000～4000 个核孔复合体。如果细胞正进行 DNA 复制，每 3min 需要从细胞质基质向核运

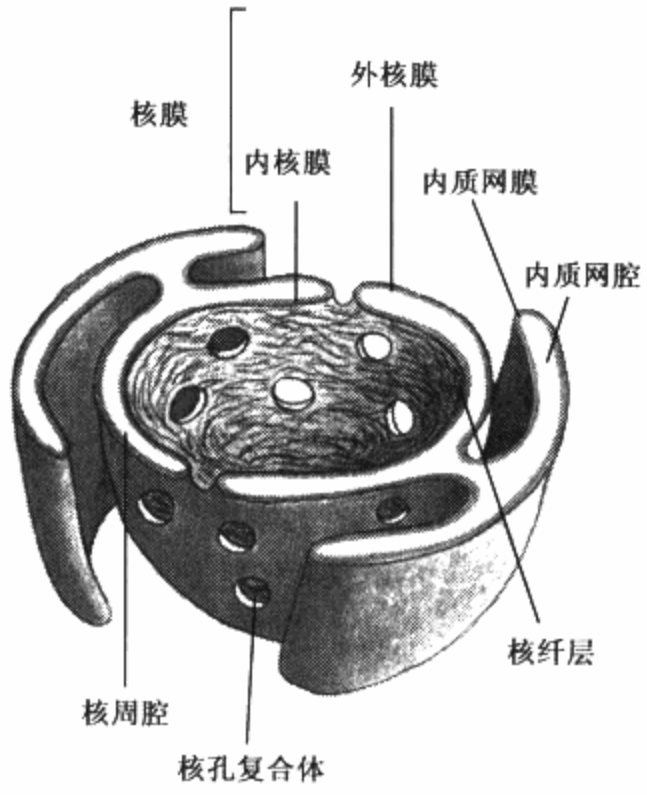


图 12-9　核被膜

双层核被膜被核孔复合体穿透并和内质网连通。图中没有显示和内质网膜胞质面以及和外核膜相结合的核糖体。核纤层是一个在内核膜下的纤维网状结构

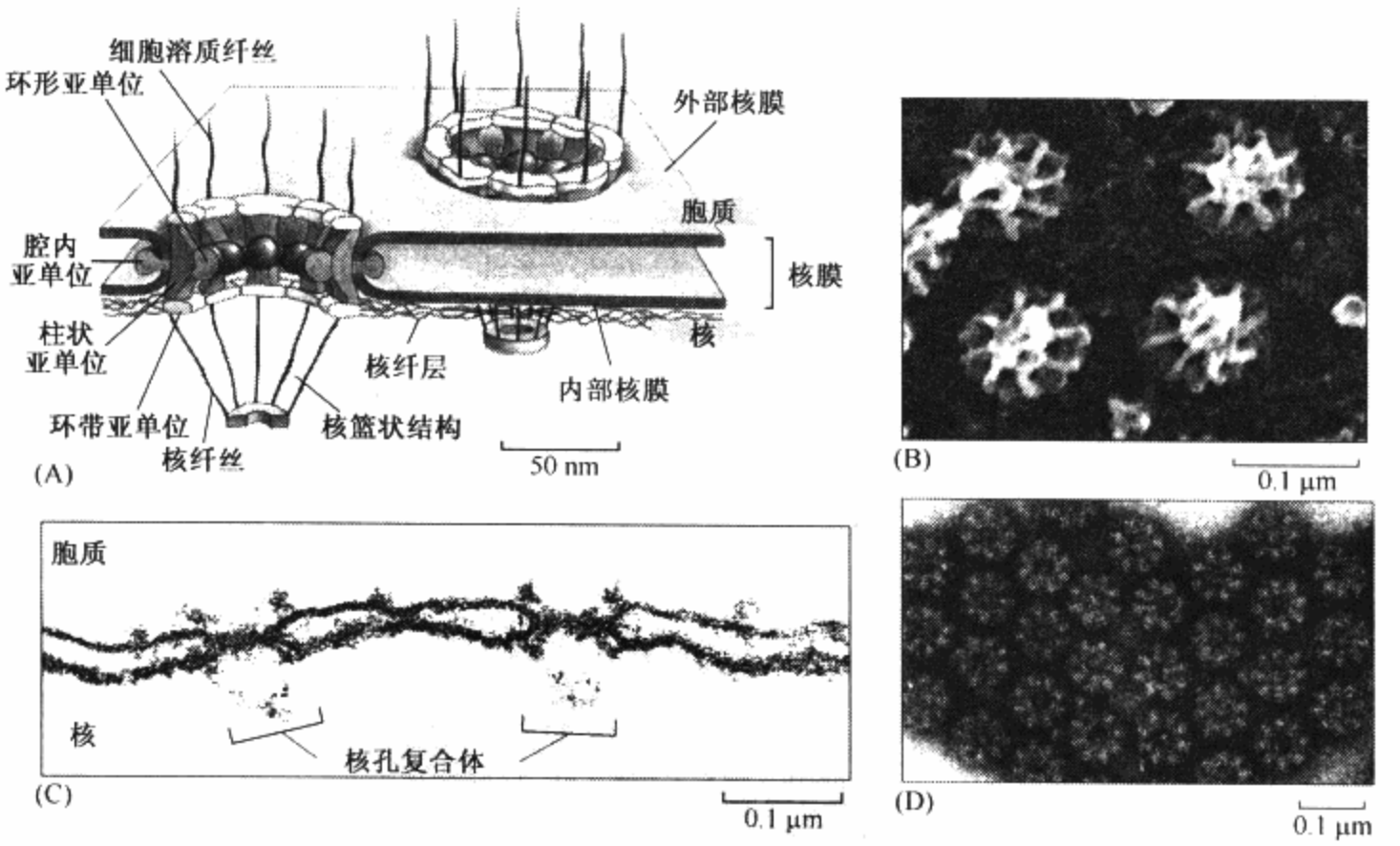


图 12-10　核孔复合体在核被膜中的结构安排

（A）核孔复合体的一小部分。在截面中，一个核孔复合体似乎具有 4 个结构区域：柱状亚单位，形成孔道的墙壁；环带亚单位，向核孔中心伸展出“辐”（未显示）；腔内亚单位，含有将复合体锚定在核膜上的跨膜蛋白；核质环，形成复合体的胞质面和核质面。此外，原纤维从复合体的胞质面和核质面伸出，在核质面，原纤维聚集形成篮状结构。利用免疫电子显微镜技术进行的定位研究显示组成核孔复合体核心的蛋白质对称地分配在核被膜中，因而核质面和胞质面看起来相同。这和组成原纤维的蛋白质不同，后者在核质面和胞质面的结构是不同的。（B）卵母细胞核被膜核质面的扫描电镜照片。（C）核孔处内核膜和外核膜的连续在这张超薄切片电镜照片中非常明显，照片显示了两个核孔复合体的旁侧结构（括号中）。（D）这张电镜照片显示了经去污剂抽提除去膜后负染核孔复合体的正面图像

入大约 106 个组蛋白分子来包裹新合成的 DNA 形成染色质，则每个核孔复合体每分钟要运入 100 个组蛋白分子。如果细胞处于快速生长期，每个核孔复合体每分钟需要从细胞核中运出大约 6 个新装配的核糖体大亚基和小亚基至细胞质基质中以供使用。这些只占通过核孔复合体的流量的一小部分。

每一核孔复合体含有一到多个开放的水通道，小的水溶性分子可通过其进行被动的自由扩散。将被标记的不同大小的水溶性分子注入细胞质基质，然后测定它们扩散进入细胞核的速率，以此来决定通道的有效大小。小分子（5000kDa 以下）的自由扩散速度非常快以至于核被膜对于它们来说是完全通透的。17 000kDa的蛋白质需要 2min 才能在细胞质基质和细胞核之间达到平衡，而大于60 000kDa的蛋白质则几乎不能进入细胞核。对这些数据的定量分析说明核孔复合体为自由扩散提供了一个长为 15nm，直径为 9nm，充满水的柱形通道，这样的通道只占核孔复合体总体积的很小一部分（图 12-11）。

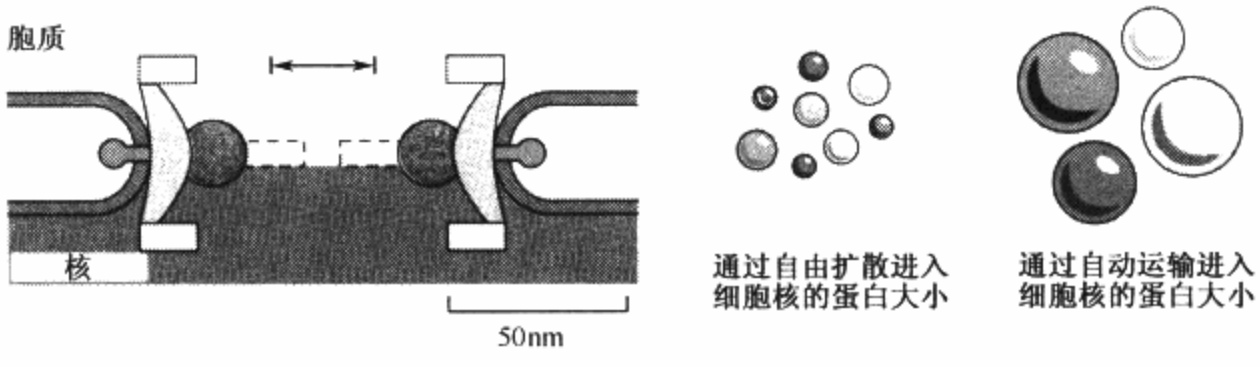


图 12-11　通过核孔复合体的可能的自由扩散途径

图中插入核孔的假想圈（灰色）将开放的通道的大小限定为 9nm，这是通过扩散测定估计的孔道大小。9nm 的直径要比电镜照片中核孔复合体所清晰显示的中心开放通道小很多，同时也比主动转运中孔道的开放程度小，因为主动转运中核孔可以膨胀到以使直径达 26nm 的颗粒通过（箭头）。因此，在准备电镜样品的过程中可能有一些核孔成分丢失了，这些丢失的成分可能限制了通过核孔的自由扩散。这些成分可能形成了一个环（或塞），它们的开合使得具有分选信号的大颗粒在主动转运过程中进入细胞核（下文讨论）。虽然在一些处理条件下可以看到塞结构，但它们究竟是核孔复合体的成分还是正在通过核孔的物质尚不清楚。计算机三维重建显示允许自由扩散的通道可能靠近核孔复合体的周边，在柱状亚单位之间而不是在核孔中央［图 12-10(A)］；这意味着被动扩散和主动转运发生在核孔复合体的不同区域

因为很多细胞蛋白质都大得难以以自由扩散的方式通过核孔复合体，所以核被膜使细胞核和细胞质基质含有不同的蛋白质成分。以成熟的细胞质核糖体为例，其直径大约 30nm，因而不能以扩散的方式通过孔径为 9nm 的通道进入细胞核；这种隔离确保蛋白质的合成仅在细胞质基质中进行。那么细胞核如何输出新生产的核糖体亚基或输入像 DNA 或 RNA 这样亚基分子质量达到100 000～200 000kDa的大分子呢？正如下文所述，上述大分子与其他蛋白质和 RNA 分子结合特异性蛋白质受体，后者运送其通过核孔复合体主动输入细胞核。

核定位信号引导核内蛋白进入细胞核

若通过实验手段将蛋白质从细胞核内抽提出来再重新导入细胞质基质中（如通过实验在细胞膜上诱导孔形成），即使是很大的蛋白质也会高效地重新聚集于细胞核中。这种选择性核运入过程是**核定位信号（NLS）**作用的结果，NLS 仅存在于核内蛋白上。通过重组 DNA 技术，已经定义了许多核内蛋白的信号（图 12-12）。如前所述，核区定位信号可以是信号序列或信号斑。大多数核内蛋白的 NLS 由 1 或 2 个富含正电荷氨基酸（赖氨酸和精氨酸）的短序列构成（表 12-3），但不同核内蛋白的精确序列有所区别。核内其他蛋白具有不同的信号，其中一些尚未加以鉴定。

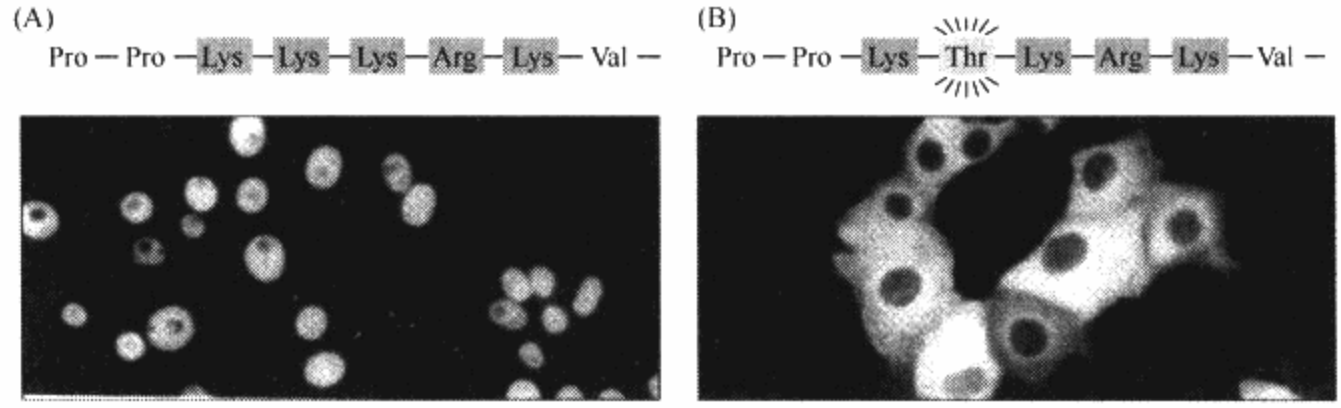


图 12-12　核定位信号的功能

免疫荧光照片显示含有或缺乏核定位信号短肽的 SV40 病毒 T-抗原在细胞中的位置。（A）正常的 T-抗原蛋白含有一个图中标注出的富赖氨酸短肽，它被转运至细胞核发挥作用，如利用抗 T-抗原的抗体进行的免疫荧光染色所示。（B）核定位信号被改变（一个苏氨酸残基代替一个赖氨酸残基）的 T-抗原驻留在细胞质基质中

已鉴定的这些信号可以定位于氨基酸序列的任何位置并在蛋白质表面形成环或斑。甚至当其以短肽链的形式连接细胞质基质蛋白质表面的赖氨酸侧链时，很多核定位信号仍能发挥功能。这说明信号在核蛋白氨基酸序列内的具体位置并不重要。

通过包被核定位信号的金微粒可以直接观察到核蛋白的核孔复合体转运：先将微粒注入细胞质基质，再用电子显微镜追踪这些微粒的路径（图 12-13）。对不同大小金微粒的研究表明，在转运过程中核孔复合体通道开口直径可以膨胀至大约 26nm。核孔复合体中心的一个装置似乎行使着“紧身圈”的功能，其开放程度恰好允许被转运物通过（图 12-11）。然而，门控机制的分子基础仍尚未揭开。

经过核孔复合体的大分子转运机制和跨其他细胞器膜的蛋白质转运机制有根本的区别，前者通过大的水通道完成而后者通过跨越一至多层类脂双层膜的蛋白质转运体实现。由于这一差别，核蛋白可以完全折叠的构象通过核孔复合体。相反地，新合成的核糖体亚基在组装完成后输出细胞核。相比之下，蛋白质转运至其他大多数细胞器时必须处于完全未折叠的状态，这将会在下文中进一步讨论。从电子显微镜观察可知，一些非常大的颗粒在穿过核孔复合体时会受到一定限制，以至于它们必须从孔道中挤过。这说明至少一些物质在通过核孔复合体时必须适当改变自身结构。如第 6 章所述，此过程已通过对大型 mRNA 分子输出的研究被更好的认识（图 6-39）。

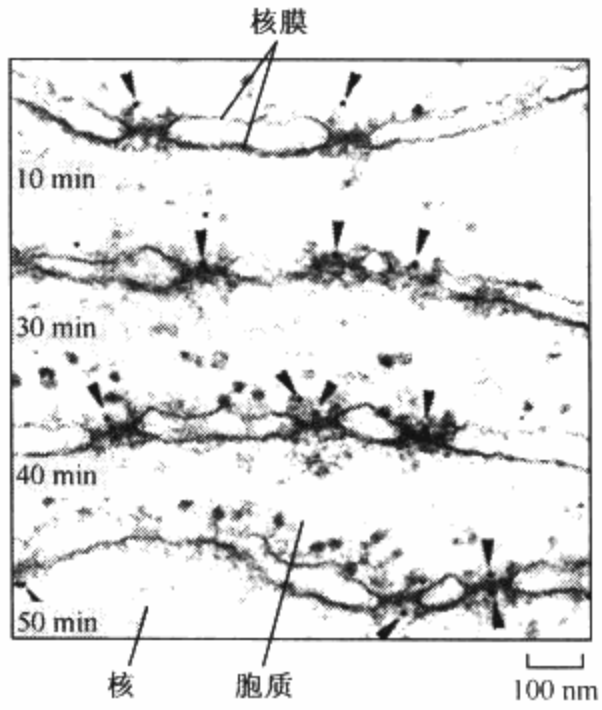


图 12-13　使核孔复合体主动转运可视化

这一组电镜照片显示被含有核定位信号的肽链包被的胶体金微粒（箭头）通过核孔复合体进入细胞核。金微粒被注入活体细胞，然后在注入后的不同时间将细胞固定，准备电镜观察。在最初（10min），金微粒接近核孔复合体的胞质原纤维，进而转移至核孔复合体的中心。此处它们先出现在核被膜胞质面上（30min 和 40min）然后出现在核质面上（50min）。这些金微粒的直径要比核孔复合体的扩散通道大得多，说明核孔被诱导变大以允许主动转运物的通过

细胞核输入受体结合核定位信号和核孔蛋白

大多数核定位信号必须被**细胞核输入受体**识别后才能引发核输入，这些核输入受体由一

个相关基因家族编码。家族中每个成员编码一种细胞核输入受体，每种输入受体专一转运一组具有相似核定位信号结构的核内蛋白［图 12-14(A)］。

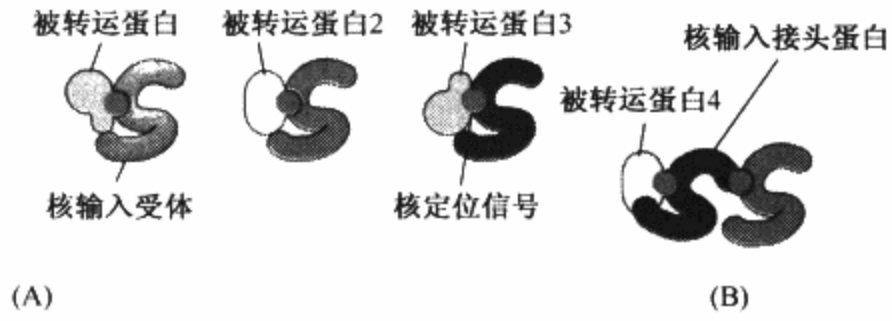


图 12-14　细胞核输入受体

（A）很多细胞核输入受体既和核孔蛋白结合又和被转运蛋白上的核定位信号结合。示例中的被转运蛋白 1、2、3 含有不同的核定位信号，导致它们结合不同的核输入受体。（B）被转运蛋白 4 需要一个接头蛋白结合它的核输入受体。接头蛋白在结构上和核输入受体相关并且识别被转运蛋白的核定位信号。接头蛋白也含有可以结合一个输入受体的核定位信号

输入受体是可溶性的细胞质基质蛋白，它结合待转运蛋白的核定位信号和核孔蛋白，其中一些核孔蛋白形成触须状的原纤维，从核孔复合体的边缘伸展至细胞质基质中。原纤维和其他核孔蛋白具有大量短氨基酸重复序列，这些重复序列因为含有苯丙氨酸和甘氨酸而被称为 FG 重复序列（用氨基酸的单字母缩写命名；讨论见第 5 章）。FG 重复序列是输入受体的结合位点。它们排列在通过核孔复合体的路径上，被输入受体和转运蛋白的复合体识别。当这些蛋白复合体通过对沿途相邻重复序列的结合、分离、再结合进入细胞核后，输入受体即和转运蛋白分离再返回细胞质基质。

核输入受体并不总是和核内蛋白直接结合，一些情况下，输入受体和转运蛋白的核定位信号的结合需要附加的接头蛋白质连接。令人吃惊的是，接头蛋白在结构上和核内运蛋白相关，说明这两种蛋白质具有共同的进化起源。输入受体和接头蛋白的结合使细胞能够识别多样的核定位信号成员。

细胞核输出和细胞核输入程序一样，方向相反

新核糖体亚基和 RNA 等大分子的细胞核输出同样经过核孔复合体并依赖于选择性的运输系统。运输系统依靠待输出大分子上的**核外运信号**以及与其互补的**核外运受体**。这些受体同时结合输出信号和核孔蛋白，引导输出蛋白经过核孔复合体进入细胞质基质。

核输出受体在结构上和核输入受体相似，它们都由同一**核转运受体**基因家族编码。在酵母中有 14 个基因参与编码这个家族的成员，而在动物细胞则要多得多。仅从氨基酸序列上经常难以区分此家族某一特定的成员是核输入受体还是核输出受体，因而很容易理解输入和输出转运系统是以相似的方式，相反的方向运转的：输入受体在细胞质基质结合待转运分子，将它们释放至细胞核，在被输出核进入细胞质基质以供再次使用，而输出受体的功能恰好相反。

如果将图 12-13 实验中使用的金微粒包被小 RNA 分子（tRNA 或 5S 核糖体 RNA），再注射进入培养细胞的细胞核，RNA 会迅速通过核孔复合体被转运至细胞质基质中。用两种大小的金微粒，一种用 RNA 包被并注入细胞核，而另一种用核定位信号包被并注入细胞质基质，结果显示每个核孔复合体都可进行双向转运。但是核孔复合体调节大分子双向流动以避免拥塞和对撞发生的机制仍不得而知。

Ran GTP 酶促使核孔复合体定向转运的发生

经过核孔复合体的核内蛋白输入过程在细胞核内累积了特定的蛋白质，这使细胞的有序性相应增加，因此是一个耗能过程（讨论见第 2 章）。消耗的能量由单体 GTP 酶 **Ran** 水解 GTP 提供。细胞质基质和细胞核中都发现存在 Ran，它是核输入和核输出系统必不可少的成分。

和其他 GTP 酶（GTPase）一样，Ran 是一种具有两种构象的分子开关，其构象形式依赖于 GDP 或 GTP 的结合（讨论见第 3 章）。构象间的转换是由两种 Ran 特异性调节蛋白引起的：细胞质基质 GTP 酶活化蛋白（GAP）可引发 GTP 水解，导致 Ran-GTP 向 Ran-GDP 的转化；核鸟嘌呤交换因子（GEF）促使 GDP 和 GTP 的交换，导致 Ran-GDP 向 Ran-GTP 的转化。由于 Ran-GAP 存在于细胞质基质中而 Ran-GEF 存在于细胞核中，因此细胞质基质中主要含有 Ran-GDP，细胞核中主要含有 Ran-GTP（图 12-15）。

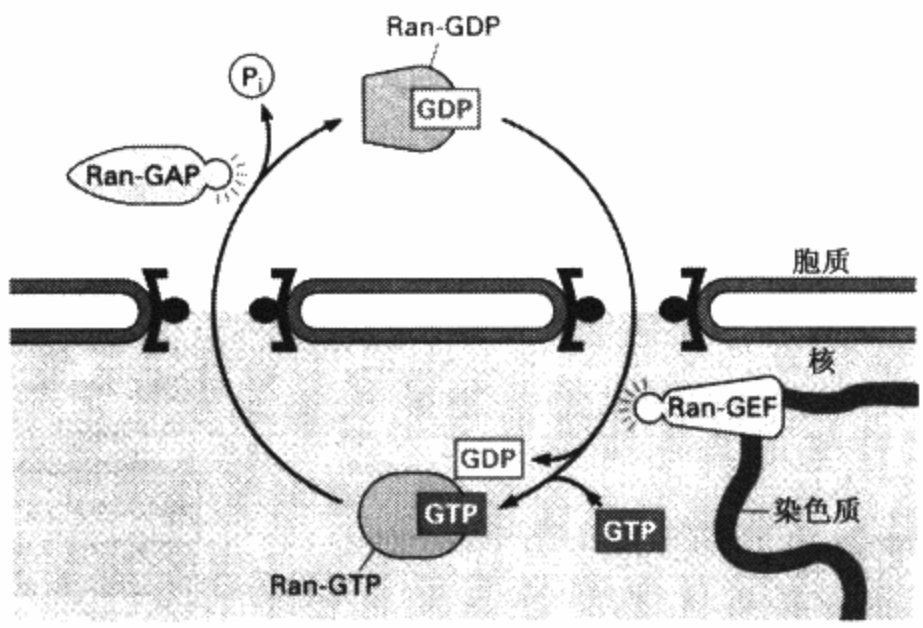


图 12-15　Ran-GDP 和 Ran-GTP 的空间划分

两种 Ran 调节蛋白的定位决定了 Ran-GDP 定位于细胞质基质而 Ran-GTP 定位于细胞核：Ran-GTP 酶活化蛋白（Ran-GAP）定位于细胞质基质，而核鸟嘌呤交换因子（Ran-GEF）由于结合在染色质上因而只存在于细胞核中。称为 Ran 结合蛋白（此处省略以使图像更清晰）的另一种蛋白质和 Ran-GAP 协同作用引发 GTP 水解

Ran 两种构象的浓度梯度驱动核转运朝正确的方向发生（图 12-16）。只有当核输入受体装载了适当的转运物时，才能和核孔复合体细胞质面的 FG-重复序列发生对接，然后输入受体及其结合的待转运蛋白沿 FG 重复序列排成的路径运动至核孔复合体核质面，再被此处的 Ran-GTP 结合，引发输入受体和转运物的分离（图 12-17）。通过细胞质基质中依赖于转运物装载的输入受体和 FG 重复序列的结合，和细胞核内依赖于 Ran-GTP 的转运物解离，Ran-GTP 的细胞核定位规定了转运的方向。

在细胞核中解离了转运物后，结合了 Ran-GTP 的空载输入受体经过核孔复合体回到细胞质基质。此处，两种细胞质基质蛋白——Ran 结合蛋白和 Ran-GAP 共同作用将 Ran-GTP 转变为 Ran-GDP。Ran 结合蛋白置换输入蛋白，和 Ran-GTP 结合，这使得 Ran-GAP 可以引发 Ran 水解结合的 GTP。然后，Ran-GDP 从 Ran 结合蛋白上解离下来并重新被输入到核中，完成整个转运周期。

核输出过程的机制十分相似，不同的是核中 Ran-GTP 促使转运物和输出受体的结合以及

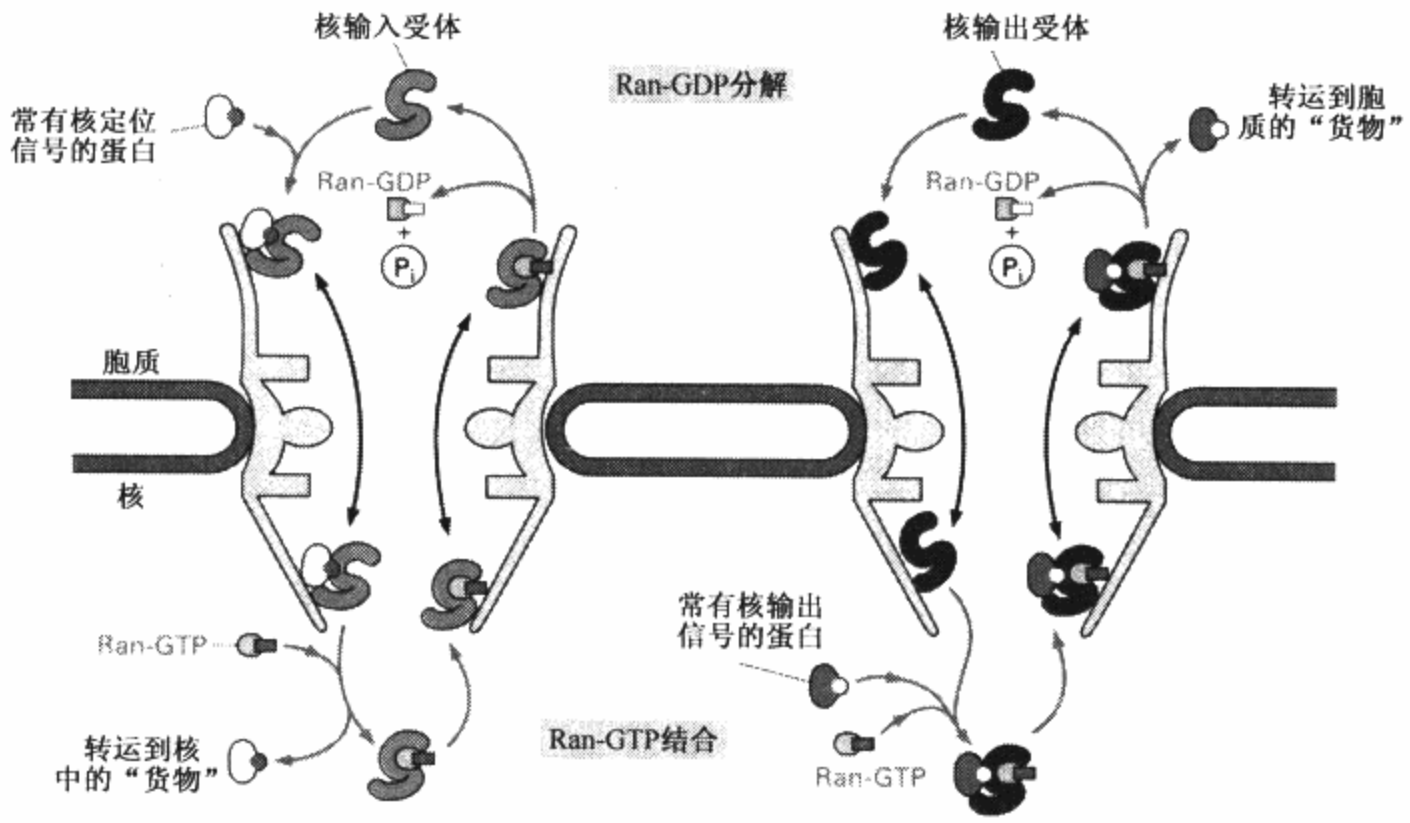


图 12-16　Ran 引发的 GTP 水解导致核定向运输的模型

核孔蛋白上 FG 重复序列引导装载了被转运蛋白的核运输受体通过核孔复合体的运动。Ran-GTP 在细胞核而 Ran-GDP 在细胞质基质的不同区域划分为核输入（左）和核输出（右）定向（红箭头）。核孔复合体胞质面上的 Ran-GAP 和 Ran 结合蛋白介导 GTP 水解产生 Ran-GDP

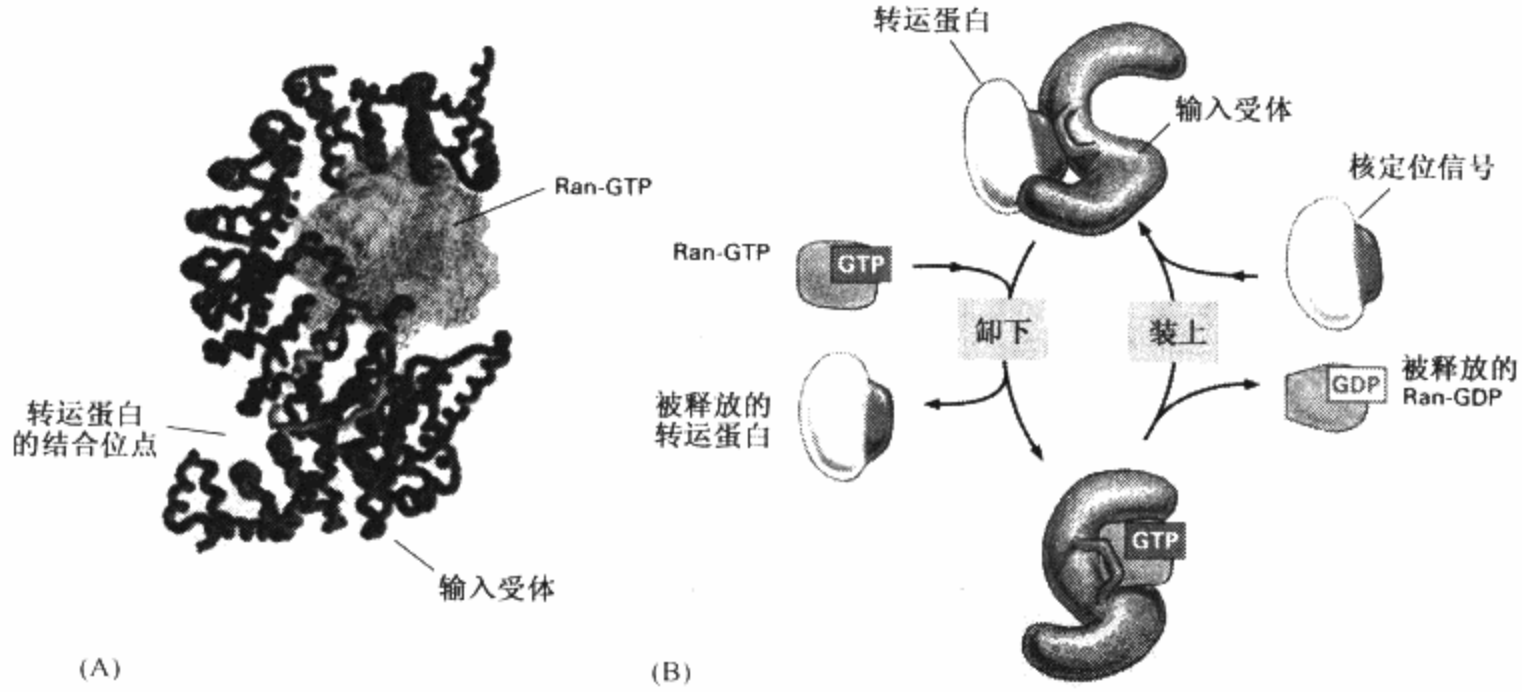


图 12-17　Ran-GTP 结合核输入受体后引发被转运蛋白释放的模型

（A）核运输受体由构成拱形或涡形螺旋的重复 α 螺旋基序组成，其组成结构取决于受体或接头蛋白的特异性。被转运蛋白和 Ran-GTP 结合拱形内面的不同区域。在核输入蛋白和 Ran-GTP 的共结晶中，受体部分的保守环（红色）被 Ran-GTP 掩盖，而这个保守环在未结合 Ran 的受体中对受体结合信号序列发挥了重要作用。（B）核输入受体在细胞质基质中的装载和在细胞核中的卸载周期

装载的受体和核孔复合体核质面的结合。一旦进入胞质面，Ran 即和 Ran-GAP 及 Ran 结合蛋白相遇，并水解自身结合的 GTP。接着输出受体将转运物和 Ran-GDP 释放入细胞质基质并从核孔复合体上解离下来，游离的输出受体再回到细胞核完成整个周期（图 12-16）。

细胞通过调控进入转运装置的途径实现细胞核和细胞质基质间转运的调节

一些蛋白质同时含有核定位信号和核输出信号，如核内结合新合成的 mRNA 的蛋白质。这些蛋白质不停的穿梭于细胞核和细胞质基质之间，其稳态位置取决于它们输入和输出的相对速度。如果输入速度超过输出速度，则这种蛋白质主要存在于细胞核。相反地，如果输出速度超过输入速度，则蛋白质主要存在于细胞质基质。因此，改变输入或输出速度可以改变这些穿梭蛋白的定位。

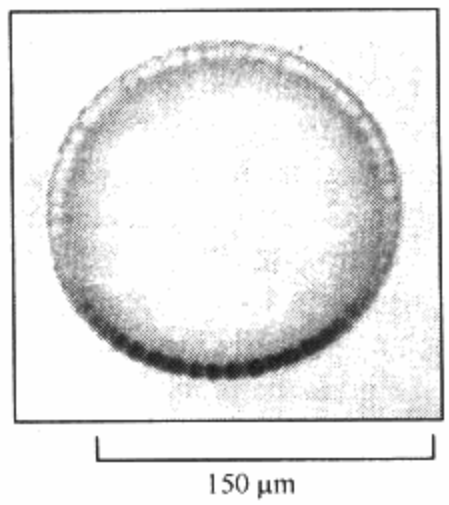


图 12-18　核运输调控果蝇胚胎的发育
截面图显示在早期的果蝇胚胎中，基因调控蛋白 dorsal 均匀的表达。但它们仅在胚胎腹面细胞的细胞核中才有活性。通过酶联抗体产生的有色物质，可以看到被特异性染色的 dorsal 蛋白

一些穿梭蛋白一刻不停的出入细胞核；而另一些的转运则在严紧控制之下。正如第 7 章讨论的，一些基因调节蛋白的活动受到控制，除非需要否则这些蛋白质会被排斥在细胞核之外（图 12-18）。大多数情况下，这种控制都是依赖核定位信号和核输出信号而完成的，因为这些信号经常可通过邻近氨基酸的磷酸化而开启或关闭（图 12-19）。

另一些基因调节蛋白和起抑制作用的细胞质基质蛋白结合，后者将调节蛋白锚定于细胞质基质中（通过和细胞骨架或特定细胞起的相互作用），或者封闭核定位信号使得它们不能和核输入受体作用。当细

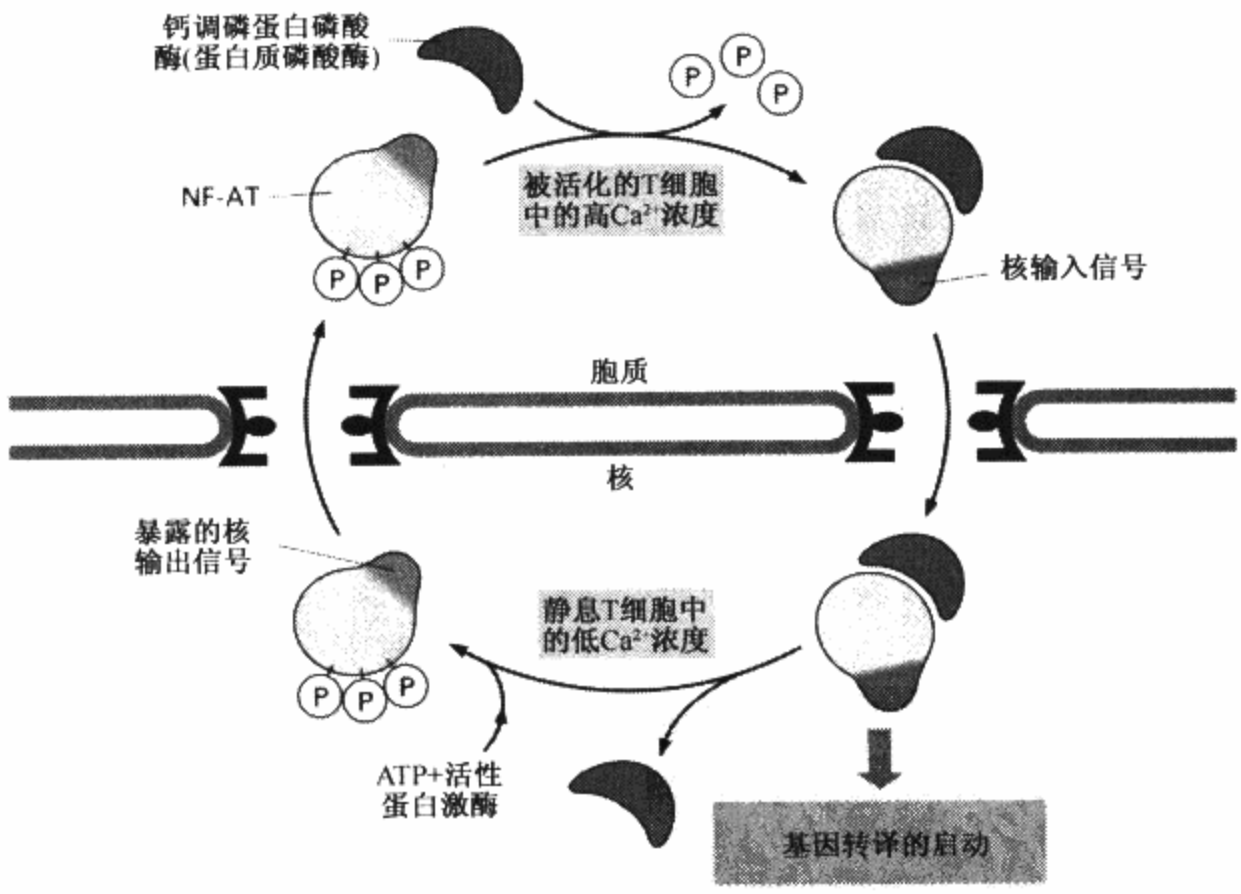


图 12-19　T 细胞激活过程中的细胞核输入调控
被激活的 T 细胞的核因子（NF-AT）是一种基因调节蛋白，在静息 T 细胞中，它们以磷酸化的状态存在于细胞质基质中。当 T 细胞被激活，细胞内 Ca^{2+} 浓度升高。在高 Ca^{2+} 情况下，蛋白质磷酸酶钙调磷蛋白磷酸酶结合 NF-AT，这引发 NF-AT 去磷酸化，暴露了一个或多个核输入信号，并且可能封闭一个核输出信号。NF-AT 和钙调磷蛋白磷酸酶的结合复合体进而被输入细胞核，在此处 NF-AT 激活正确免疫应答所需的大量细胞因子和细胞表面蛋白基因的转录。在应答结束的过程中，Ca^{2+} 的降低导致钙调磷蛋白磷酸酶的释放。NF-AT 的复磷酸化抑制核输入信号，重新暴露 NF-AT 的核输出信号，导致 NF-AT 重新回到细胞质基质中。一些效力很强的免疫抑制药物，如环孢霉素 A 和 FK506，可以抑制钙调磷蛋白磷酸酶对 NF-AT 的去磷酸化能力，因此可以防止 NF-AT 在细胞核中的累积

胞受到适当的刺激时，基因调节蛋白从细胞质基质的锚定或封闭中释放出来并被转运进入细胞核。一个重要的例子是控制胆固醇代谢途径相关蛋白质表达的潜在基因调节蛋白。这种调节蛋白以无活性的形式合成，以跨膜蛋白的形式储存在内质网上。当缺乏胆固醇时，细胞会激活特定的蛋白酶作用于这些无活性的蛋白质，释放它的细胞质基质结构域。进而这一结构域被输入细胞核，激活胆固醇输入及合成必需基因的转录。

细胞调控 RNA 核输出的机制和输入机制很相似。mRNA 在转录和剪接的过程中会和一些蛋白质结合，后者具有核输出信号，可被输出受体识别从而介导 RNA 通过核孔复合体输出细胞核。进入细胞质基质后，包被了 RNA 的蛋白质会立刻剥落 RNA 并迅速回到细胞核中。其他 RNA，如 snRNA 和 tRNA，是通过不同种类的核输出受体输出核的。

未完成修饰处理的前 mRNA 被保留在细胞核中，锚定在核转录和剪接装置上。这种装置只在 RNA 分子完成修饰后才将其释放。酵母的遗传学研究表明：不能正确结合剪接装置的突变前 mRNA 在未完成剪接的情况下就被不正确的输出细胞核。

核被膜在有丝分裂过程中分解

核纤层是由称为核纤层蛋白的一种互联蛋白质亚基构成的网状结构。核纤层蛋白是一类特殊的中间纤维蛋白（讨论见第 16 章），聚合形成一个二维的网络（图 12-20）。核纤层通过与核孔复合体和内核膜膜内在蛋白的连接锚定在核被膜上，使其具有一定形状并且维持其稳定。核纤层也直接和染色质相互作用，后者还和内核膜的膜内在蛋白作用，这些膜蛋白和核纤层一起为 DNA 和核被膜提供结构连接。

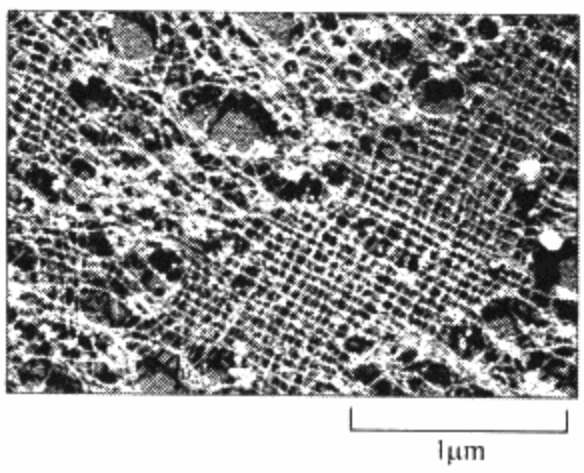


图 12-20　核纤层

电镜照片示通过冷冻干燥和金属投影制备的非洲爪蟾卵母细胞的部分核纤层。核纤层由特殊的中间纤维构成的规则网格形成

当细胞核在有丝分裂过程中解体时，核纤层解聚，这是由于在有丝分裂开始时细胞周期素依赖性蛋白激酶的激活导致的核纤层蛋白磷酸化所致（讨论见第 17 章）。同时，内核膜蛋白也被磷酸化，核孔复合体解体并分散至细胞质基质中。核被膜蛋白由于不再束缚于核孔复合体、核纤层或染色质，因而扩散至整个内质网膜。所有这些事件都导致分隔细胞核和细胞质基质的屏障崩解，分裂细胞中不再结合核膜或染色体的核内蛋白完全和细胞质基质混杂（图 12-21）。

在有丝分裂后期，内核膜蛋白和去磷酸化的核纤层蛋白重新结合染色质，核被膜在染色体的表面重新集合排列。内质网膜包裹染色体并不断融合直到一个封闭的核被膜完成重建。在这一过程中，核孔复合体也重组并开始主动输入那些具有核定位信号的蛋白质。当新的核被膜在紧靠染色体表面的位置刚刚形成时，新生的细胞核内除了最初结合在有丝分裂染色体上的蛋白质和通过核孔复合体选择性输入的蛋白质外，没有其他的蛋白质。这样，其他所有的大蛋白质分子都被排斥在新核之外。

在蛋白质转运进入细胞核后，核定位信号并不被切下。这很可能是由于每一次细胞分裂后核内蛋白都需要被重新输入细胞核。相比之下，当一个蛋白质分子输入到其他膜包被的细胞器后，它会世代停留于此细胞器中而不再需要易位，所以这些分子的信号序列往往在蛋白质转运后被移除。

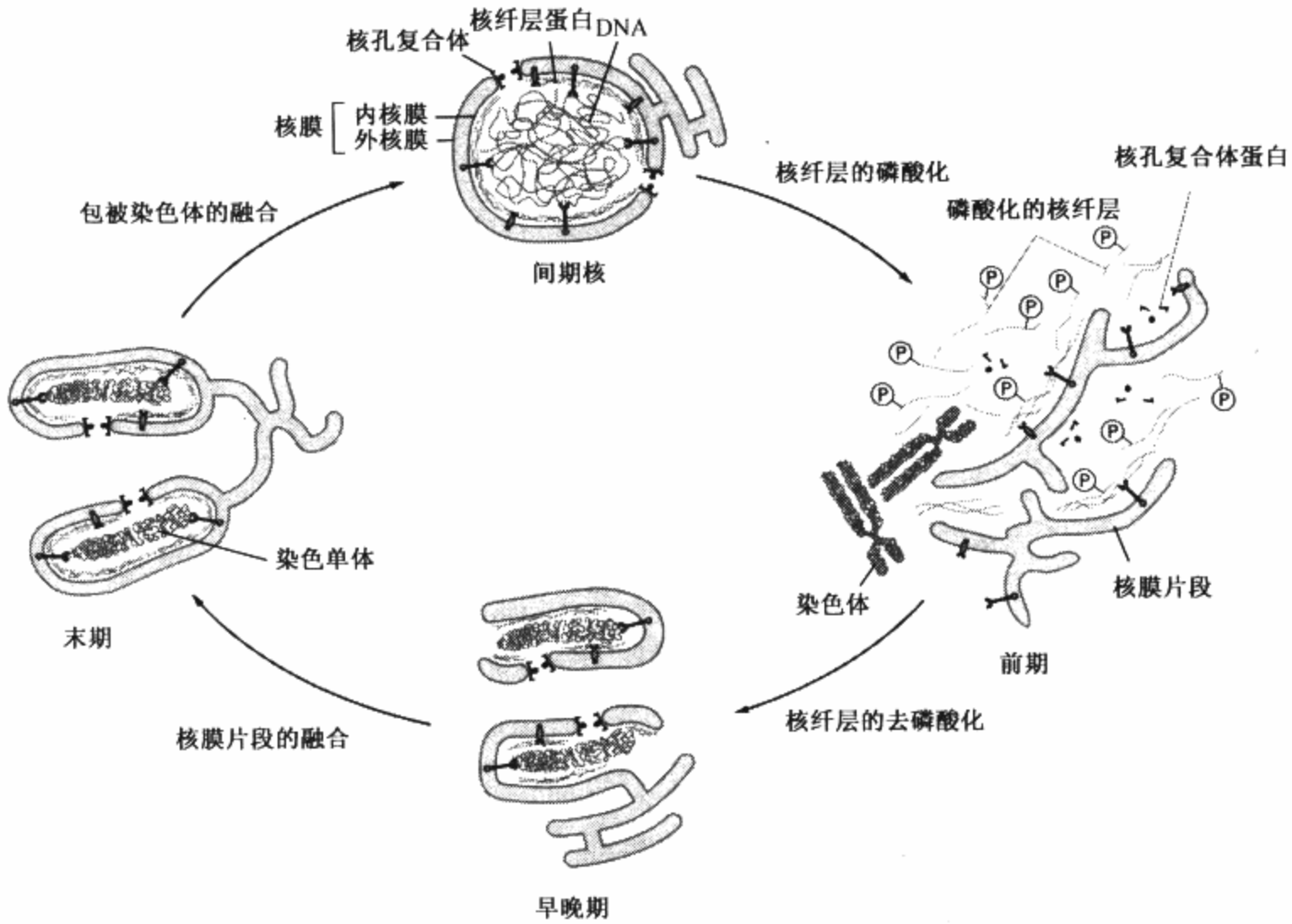


图 12-21　有丝分裂期间核被膜的解聚和重建

核纤层的磷酸化被认为引发了核纤层的解体，进而引起核被膜的降解。而核纤层的去磷酸化则逆转了这一过程

小结

核被膜有内核膜和外核膜组成，外核膜和内质网膜连接，内外核膜间的间隙和内质网腔连通。产生于核的 RNA 分子和组装于核的核糖体亚基被输出至细胞质基质；而所有在核内行使功能的蛋白质都在细胞质基质中合成后输入细胞核。这种核质间普遍存在的物质运输通过核孔复合体完成，核孔复合体成为跨越核被膜的直接运输通路。

含有核定位信号的蛋白质通过核孔复合体被主动输入至核内，而含有核输出信号的 RNA 分子和新产生的核糖体亚基通过核孔复合体被主动输出至核外。包括核输入和输出受体在内的一些蛋白质持续穿梭于细胞质基质和细胞核之间。GTP 酶 Ran 为核运输提供了方向性。核内蛋白和 RNA 分子通过核孔复合体的运输可通过抑制这些分子进入转运装置的途径进行调控。由于核定位信号不被移除，因而核内蛋白可以在每次有丝分裂后核重建时被运入细胞核。

线粒体和叶绿体的蛋白质输入

正如第 14 章所述，线粒体和叶绿体是由双层膜包被的细胞器。它们利用线粒体中电子传递和氧化磷酸化以及叶绿体中光合磷酸化产生的能量，专门负责 ATP 的合成。虽然两种细胞器都含有自己的 DNA、核糖体及其他蛋白质合成的必需元件，其大部分蛋白质仍是在细胞核内编码再输入细胞质基质中合成的。每一种核编码的蛋白质都必须进入特定细胞器后才正确行使功能。

线粒体内分为两个空间：**内部嵴间隙**和**膜间隙**。这两部分空间由两层线粒体膜形成：内

膜形成广泛的内陷，即嵴，包围嵴间隙；**外核膜**和细胞质基质接触［图 12-22(A)］。叶绿体同样具有类似的两个空间，此外还额外具有被类囊体膜包围形成的类囊体腔［图 12-22(B)］。线粒体和叶绿体的每个空间内都含有一组特殊的蛋白质。

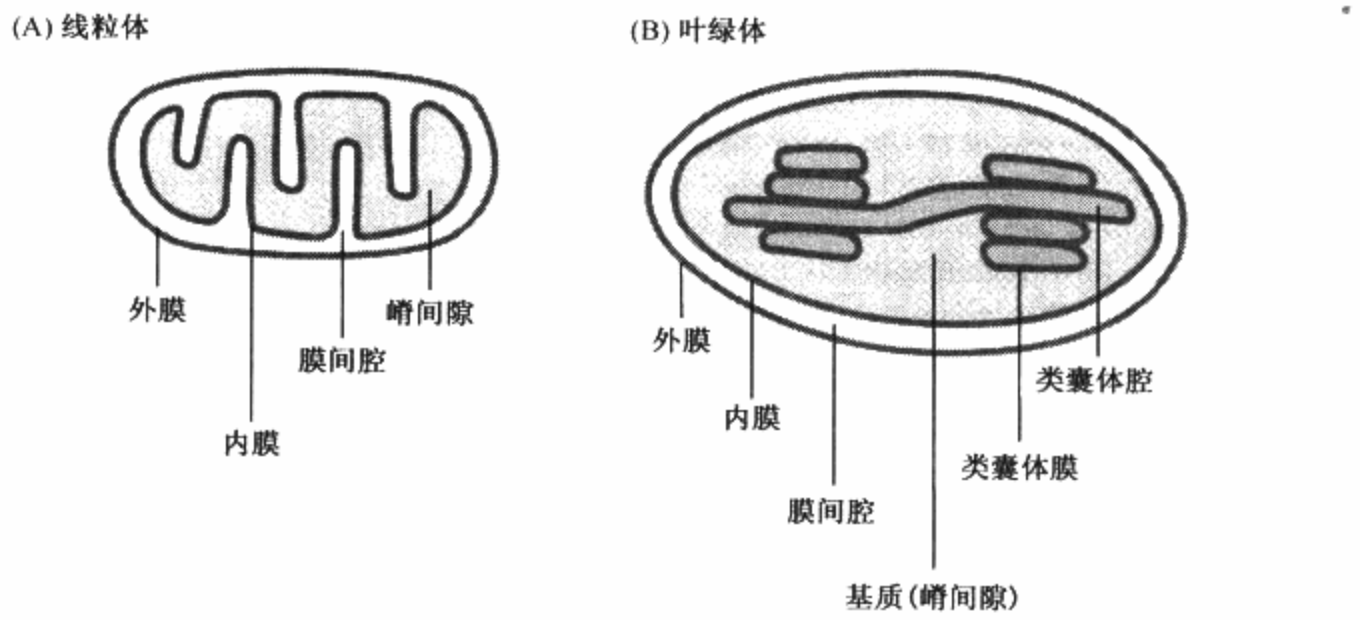


图 12-22　线粒体和叶绿体的胞内亚分隔

和线粒体嵴（A）相比，叶绿体的类囊体（B）不和内膜连接因此形成一个具有独立内部空间的包内分隔

新的线粒体和叶绿体由前已存在的相应细胞器通过生长分裂产生（讨论见第 14 章）。它们的生长主要依赖于细胞质基质蛋白质的输入。这需要蛋白质穿过多层连续膜最终进入适当的空间位置。这一过程如何发生将是本节探讨的主题。

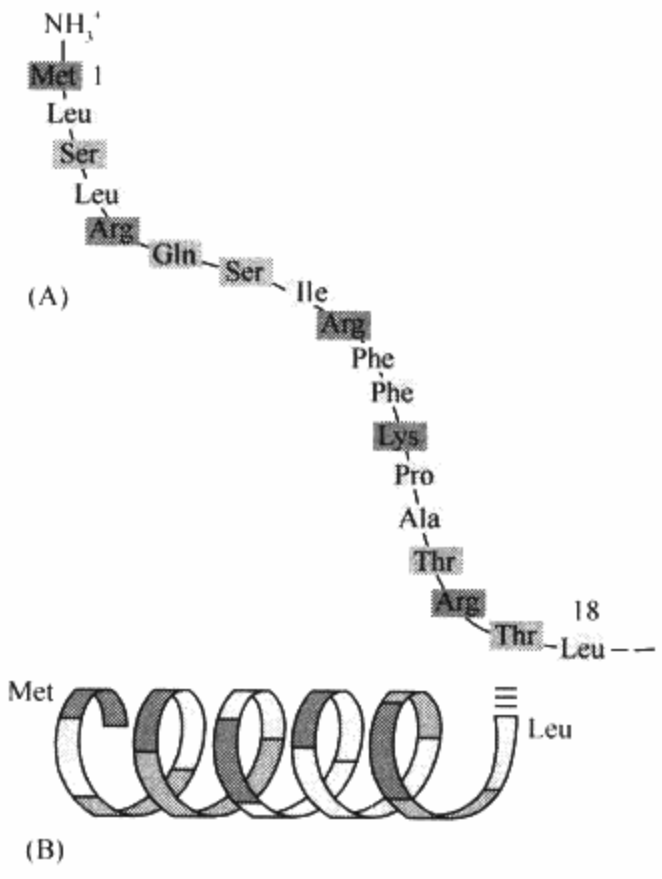


图 12-23　线粒体蛋白质输入的信号序列

细胞色素蛋白酶是定位在线粒体内膜上的一种大型多蛋白复合体，其功能是电子传递链中的末端转移酶（讨论见第 14 章）。（A）此酶的第 IV 亚基前体的前 18 个氨基酸作为信号序列，将这个亚基输入线粒体。（B）当信号序列折叠成 α 螺旋时，带正电荷的残基（红色）被聚集在螺旋的表面一侧，而非极性残基（黄色）被主要聚集在表面另一侧。线粒体基质定位序列总是形成这种两亲性的 α 螺旋，被线粒体表面特异性的受体蛋白识别

蛋白质易位进入线粒体基质依赖于信号序列和蛋白质易位体

线粒体基质蛋白被核糖体释放几秒钟至几分钟后就从细胞质基质中输入线粒体基质。所以，和下文介绍的蛋白质易位进入内质网的过程相比，线粒体蛋白质先以前体蛋白的形式在细胞质基质中完成合成后，再通过翻译后机制易位进入线粒体。大多数**线粒体前体蛋白**在 N 端具有一个信号序列，输入线粒体基质后这一序列被蛋白酶（信号肽酶）迅速切除。拥有信号序列是线粒体蛋白质输入的充分必要条件：采用基因工程技术将这些信号和任何细胞质基质蛋白质相连，可以引导这些蛋白质进入线粒体。通过序列比较和物理学研究不同基质信号序列的共同特征是倾向于折叠形成 α 螺旋，且带正电荷的残基聚集在螺旋的一侧而不带电荷的输水残基聚集在相反的一侧（图 12-23）。是这种构型而

并非精确的氨基酸序列被引发蛋白质易位的特异性蛋白质受体识别。

跨线粒体膜的**蛋白质易位**被行使易位体功能的多亚基蛋白复合体介导：**TOM 复合体**负责外膜的穿越，而两种 **TIM 复合体**——TIM23 和 TIM22 复合体负责内膜的穿越（图 12-24）。TOM 和 TIM 分别是线粒体外膜易位酶和线粒体内膜易位酶的英文名称缩写。这些复合体的一些部分是线粒体前体蛋白的受体而另一些部分则形成易位通道。TOM 复合体是所有核编码线粒体蛋白输入线粒体过程所必需的，它先将线粒体蛋白的信号序列运输至膜间隙中并帮助跨膜蛋白插入外膜。然后 TIM23 复合体将其中一部分蛋白质运入基质中，且帮助跨膜蛋白插入内膜。TIM22 介导内膜蛋白这一亚类的插入，其中包括运输 ADP、ATP 和磷酸的载体蛋白。细粒体内膜上的第 3 种蛋白质易位子被称作 **OXA 复合体**，它负责介导线粒体内合成的蛋白质插入内膜蛋白。这种复合体也可帮助最初通过 TOM 和 TIM 复合体进入线粒体细胞质基质的部分蛋白质插入内膜。

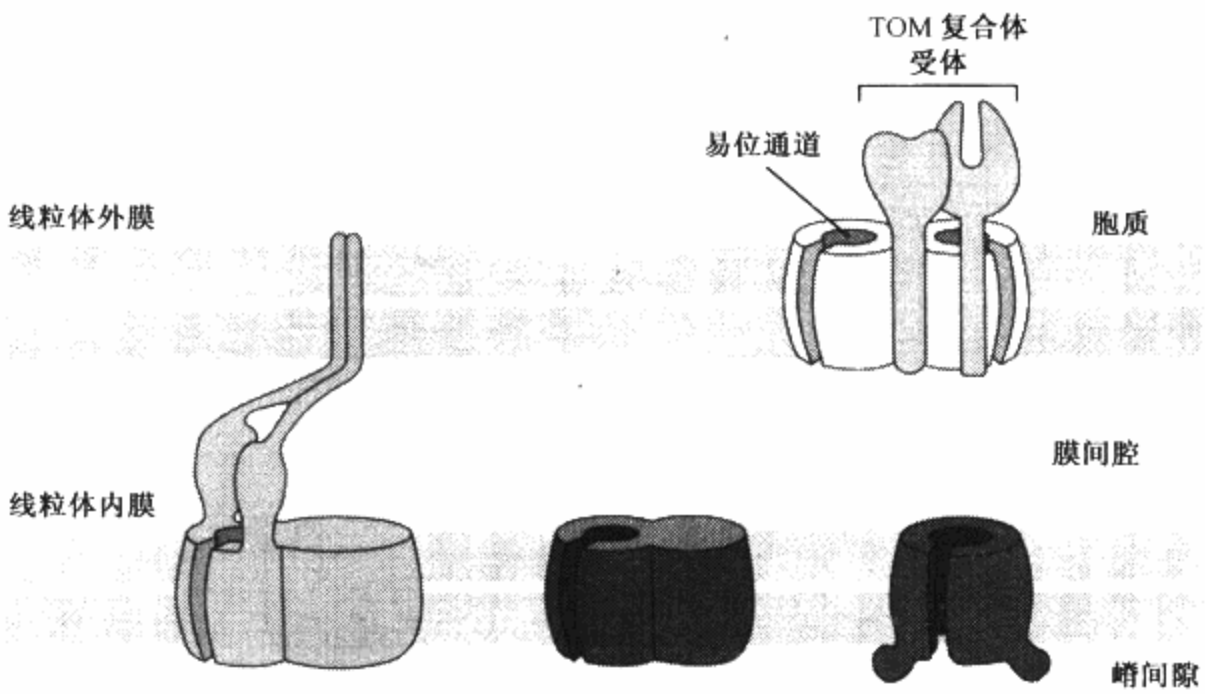


图 12-24　3 种线粒体膜上的蛋白质易位体

TOM 和 TIM 复合体以及 OXA 复合体是膜蛋白多聚体装置，负责催化蛋白质的线粒体跨膜运输。组成 TIM22 和 TIM23 复合体输入通道的蛋白质成分具有结构相似性，说明这两种 TIM 复合体有共同的进化起源。如图所示，TIM23 复合体的主要成分之一具有一个插入线粒体外膜的疏水性 α 螺旋延伸；这种复合体因为同时穿过两层膜而不同寻常

线粒体前体蛋白以伸展的多肽链形式输入线粒体

现今所了解的几乎所有有关线粒体蛋白质输入分子机制的内容都是通过分析非细胞重建运输体系得到的。先通过匀浆细胞的差速离心分离纯化出线粒体，再将其和放射性标记的线粒体前体蛋白共育。后者将被迅速而高效地吸收。通过改变共育条件，有可能建立起运输需要的生化条件。

线粒体前体蛋白在合成后并不折叠形成天然的三维结构，而是维持在伸展的状态。这需要前体蛋白和细胞质基质中其他蛋白质的相互作用：其中一些互作蛋白是属于热激蛋白 70（hsp70）家族的伴侣蛋白；另一些则是线粒体前体蛋白的专一结合蛋白，作用于它们的信号序列。这些蛋白质都能防止前体蛋白在结合线粒体外膜的 TOM 复合体前的自发聚集或折叠。作为输入过程的第一步，线粒体前体蛋白上的信号序列先被 TOM 复合体上的输入受体蛋白识

别并结合，随后细胞质基质中的互作蛋白质脱落，从信号序列开始，伸展的多肽链进入易位通道。

线粒体前体蛋白从内膜和外膜的接触点输入线粒体基质

原则上说，蛋白质可以依次穿过两层膜或一次性穿过两层膜到达线粒体基质。为了确定哪一种易位方式的可能性更大，将无细胞线粒体输入体系冷却至低温，使正处于易位过程的蛋白质停留在中间步骤。处理结果显示累积在中间步骤的蛋白质其 N 端信号序列已被基质信号肽酶切除，说明它们的 N 端已在线粒体嵴间隙中。但是，蛋白质的大部分仍能被在线粒体外部添加的蛋白酶水解（图 12-25）。这一结果说明前体蛋白可以一次性同时穿过双层膜进入线粒体基质（图 12-26）。普遍认为 TOM 复合体先运送线粒体定位信号穿过外膜，一旦到达膜间隙，定位信号就结合在 TIM 复合体上，使其通道打开。多肽链通过 TIM 复合体上的通道进入线粒体基质或插入内膜上。电子显微镜技术人员已经发现了大量内外膜紧密连接的接触位点，易位似乎多发生于这些位点上或附近。

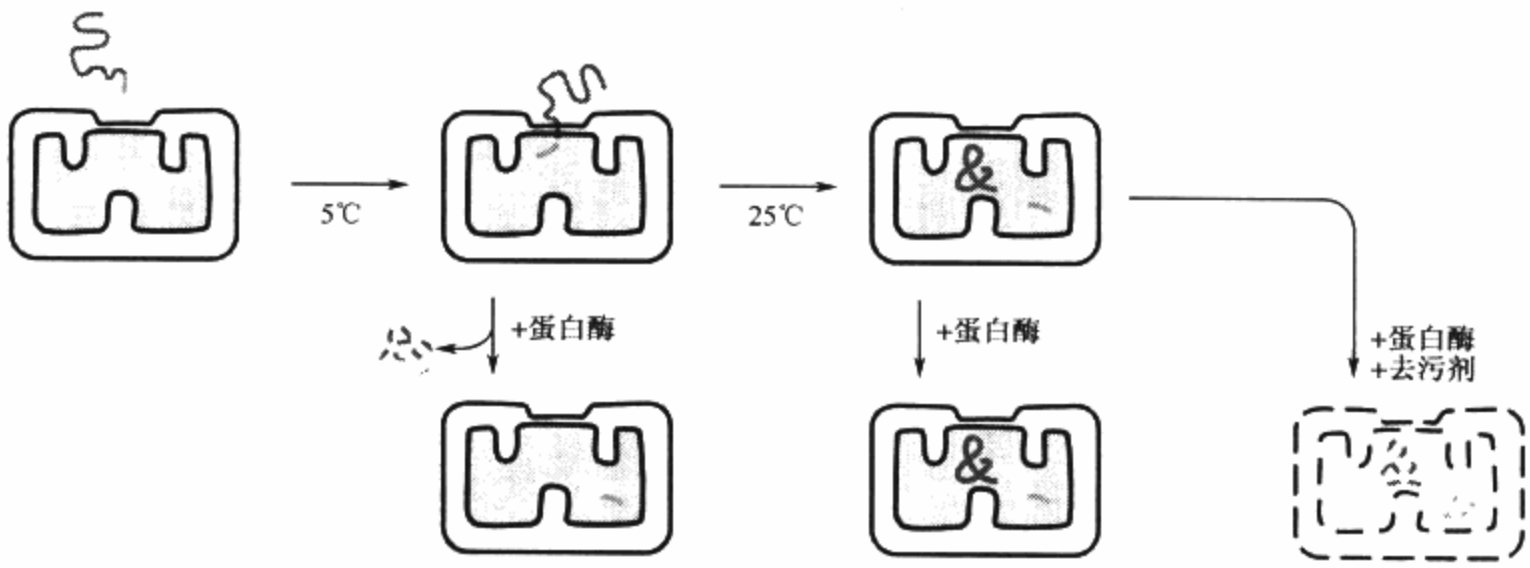


图 12-25　在易位进入线粒体基质的过程中，蛋白质瞬间同时穿过线粒体的内膜和外膜

把分离的线粒体和一种前体蛋白共培养于 5℃，前体蛋白仅被部分易位。在线粒体基质中，N 端信号（红色）被切除；但是大部分多肽链仍被保留在线粒体外，因此能被蛋白水解酶降解。当加温到 25℃时，蛋白质完全易位。一旦完全进入线粒体，多肽链就不会被外界加入的蛋白水解酶降解。作为对照组，当用去污剂破坏线粒体膜后，被输入的蛋白质又可被同样的蛋白酶处理容易地降解

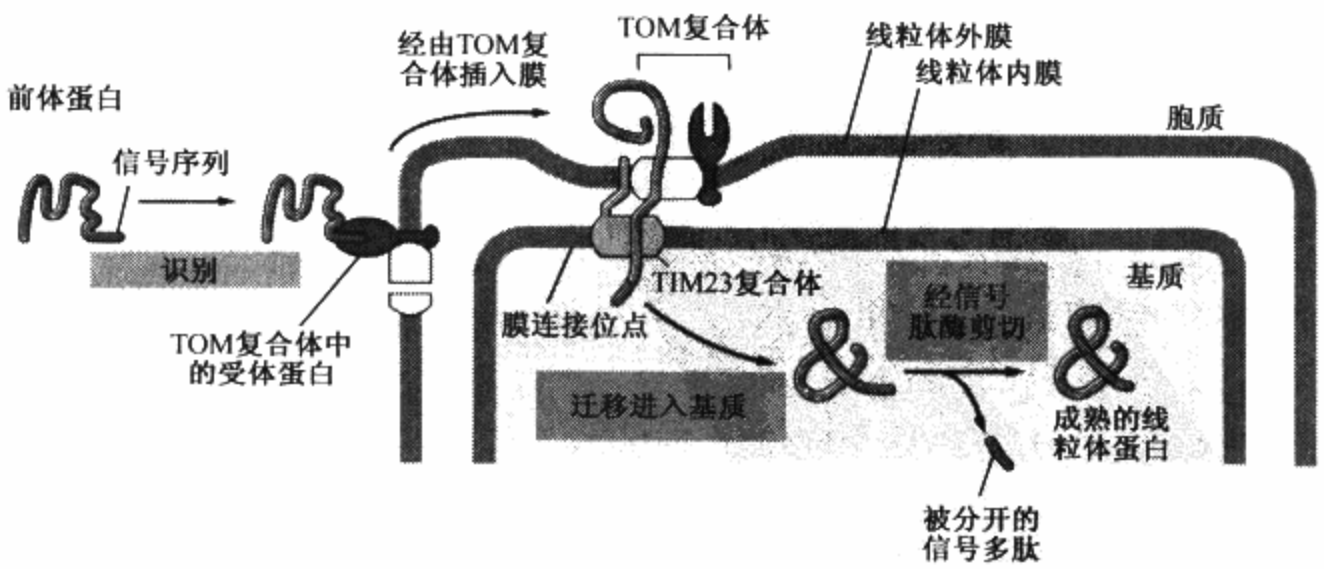


图 12-26　蛋白质的线粒体输入

前体蛋白的 N 端信号序列被 TOM 复合体的受体识别。蛋白质的易位发生在线粒体双层膜的特殊的接触位点或靠近接触位点的位置。信号序列被基质中的信号肽酶切除，形成成熟的蛋白质。切除的信号序列随即被降解（未显示）

虽然 TOM 和 TIM 复合体通常偶联使被运送的蛋白质一次性穿过双层膜，但是两种易位体可以独立工作。例如，将外膜分离，其上的 TOM 复合体仍具有运送前体蛋白的信号序列跨膜的能力。相似地，用实验方法破裂线粒体外膜，暴露在表面的 TIM23 复合体可以高效地运送前体蛋白质进入嵴间隙。但即使 TOM 和 TIM 易位体可以独立行使功能，线粒体接触位点的内外膜还是始终被穿过这两层膜的 TIM23 复合体结合在一起。

ATP 的水解和 H^+ 梯度驱动蛋白质输入线粒体

定向运输是需能过程。在多数生物系统中，能量由 ATP 水解提供。线粒体蛋白质输入由发生在两个独立场所的 ATP 水解驱动：其一是在线粒体外，另一是在线粒体基质中(图 12-27)。此外，另一能量来源也被用于定向运输:线粒体内膜两侧的 H^+ 电化学梯度。

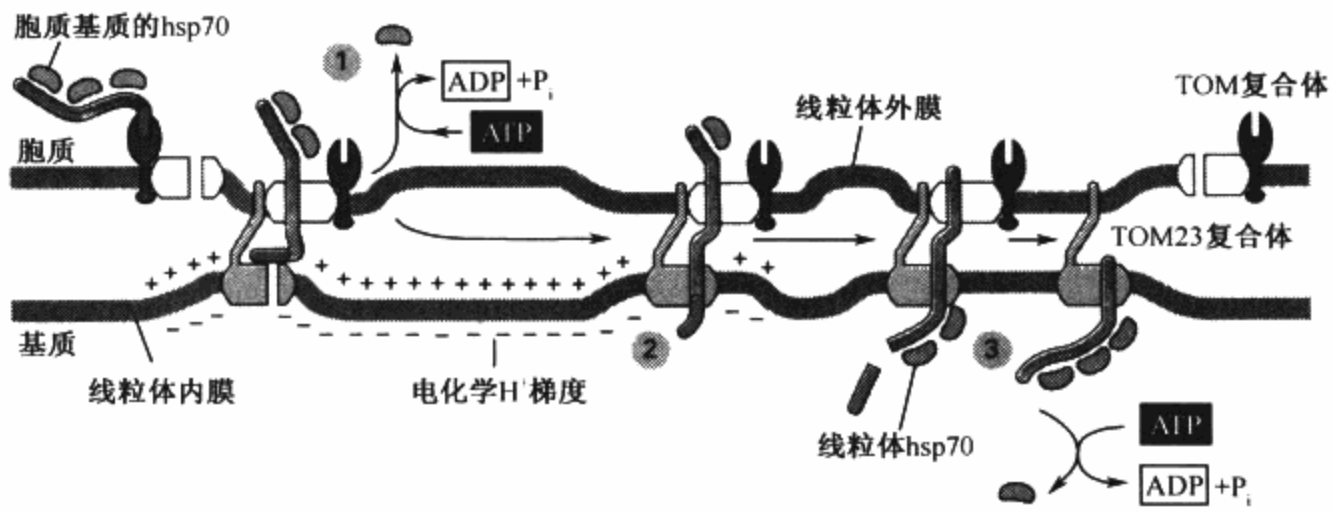


图 12-27　蛋白质输入线粒体基质过程的能量作用

①结合在被转运蛋白上的细胞质基质 hsp70 和蛋白质分离，这一环节依赖 ATP 的水解。当信号序列和紧连的多肽链部分最初插入 TOM 复合体之后，信号序列和 TIM 复合体相互作用。②然后信号序列易位进入基质，这一环节需要跨内膜的 H^+ 电化学梯度将舒展的蛋白质定位，使之瞬间同时跨过两层膜。③线粒体 hsp70 结合已经暴露于基质的多肽链区域，进而将蛋白质“拉入”基质中。然后 ATP 水解导致线粒体 hsp 从蛋白质上分离，使已经输入的蛋白质折叠

第一个需能过程发生在易位最初阶段，即结合伴侣蛋白的伸展状前体蛋白和线粒体输入受体相互作用的阶段。正如第 6 章中所述，当新合成的多肽链被热激蛋白 70 家族伴侣蛋白释放时，ATP 的水解为这一过程提供了能量。从实验角度讲，如果在加入分离的线粒体前前体蛋白能够被人工舒展开，此步就不需要热激蛋白 70 和 ATP 的参与。

当信号序列通过 TOM 复合体并结合任意一种 TIM 复合体后，通过 TIM 的进一步易位需要内膜两侧的 H^+ 电化学梯度。内膜的电子传递链驱动 H^+ 从基质中持续泵入膜间隙以维持电化学梯度。相比之下，线粒体外膜则和革兰氏阴性菌相似（图 11-17），具有一种称为膜孔蛋白的孔生成蛋白质，允许无机离子和代谢产物（不包括大部分蛋白质）自由通过，因而在外膜的两侧无法建立起离子梯度。横跨内膜的 H^+ 电化学梯度能量不但用于驱动细胞大部分 ATP 的合成；还用于驱动定位信号通过 TIM 复合体的易位。虽然尚不知 H^+ 梯度提供能量的精确机制，但有可能是这一梯度的电势部分（膜电势，图 14-13）促使带正电荷的信号序列通过电泳进入线粒体基质。

线粒体嵴间隙中的热激蛋白 70 伴侣蛋白在易位过程中也发挥作用，它们是输入过程中消耗 ATP 的第 3 个环节，将在下文中讨论。

线粒体 hsp70 催化的 ATP 水解重复循环使输入过程得以完成

我们知道线粒体 hsp70 对于输入过程至关重要，因为此蛋白质突变的线粒体不能输入前体蛋白。和细胞质基质中的同类蛋白质相似，线粒体 hsp70 对伸展的多肽链具有较高的亲和能力，因此当输入的蛋白质刚从易位体进入基质时，就被其紧密地结合。随后 hsp70 释放蛋白质，这是一个依赖 ATP 的环节。在输入蛋白刚刚插入 TIM23 复合体后，正是这一能量驱使的重复结合释放环节为整个蛋白质输入的完成提供了最终驱动力。

已有 2 个模型来解释通过线粒体 hsp70 的线粒体水解如何驱动蛋白质的输入。在这两个模型中，hsp70 都和 TIM23 复合体紧密结合，后者将前者置于刚刚进入基质的易位多肽链上。在热棘齿模型（thermal ratchet model）中［图 12-28(A)］，多肽链由于热运动可以在 TIM23 易位通道中来回滑动，每一次向基质内滑出足够的长度，接着一个 hsp70 分子结合其上防止滑出的部分退回易位体，保证了蛋白质的定向运动。这样一个接一个的多个 hsp70 的结合将多肽链易位进入基质。在横桥锁止扣模型（cross-bridge ratchet model）中［图 12-28(B)］，hsp70 结合了进入基质的多肽链后在 ATP 的水解下发生构象转变，主动将一段多肽链拉入基质。然后一个新的 hsp70 分子再结合在刚被拉入的片段上重复上述过程。因此，在上述两个模型中，hsp70 都在发挥锁止扣的功能以防止多肽链的回退。

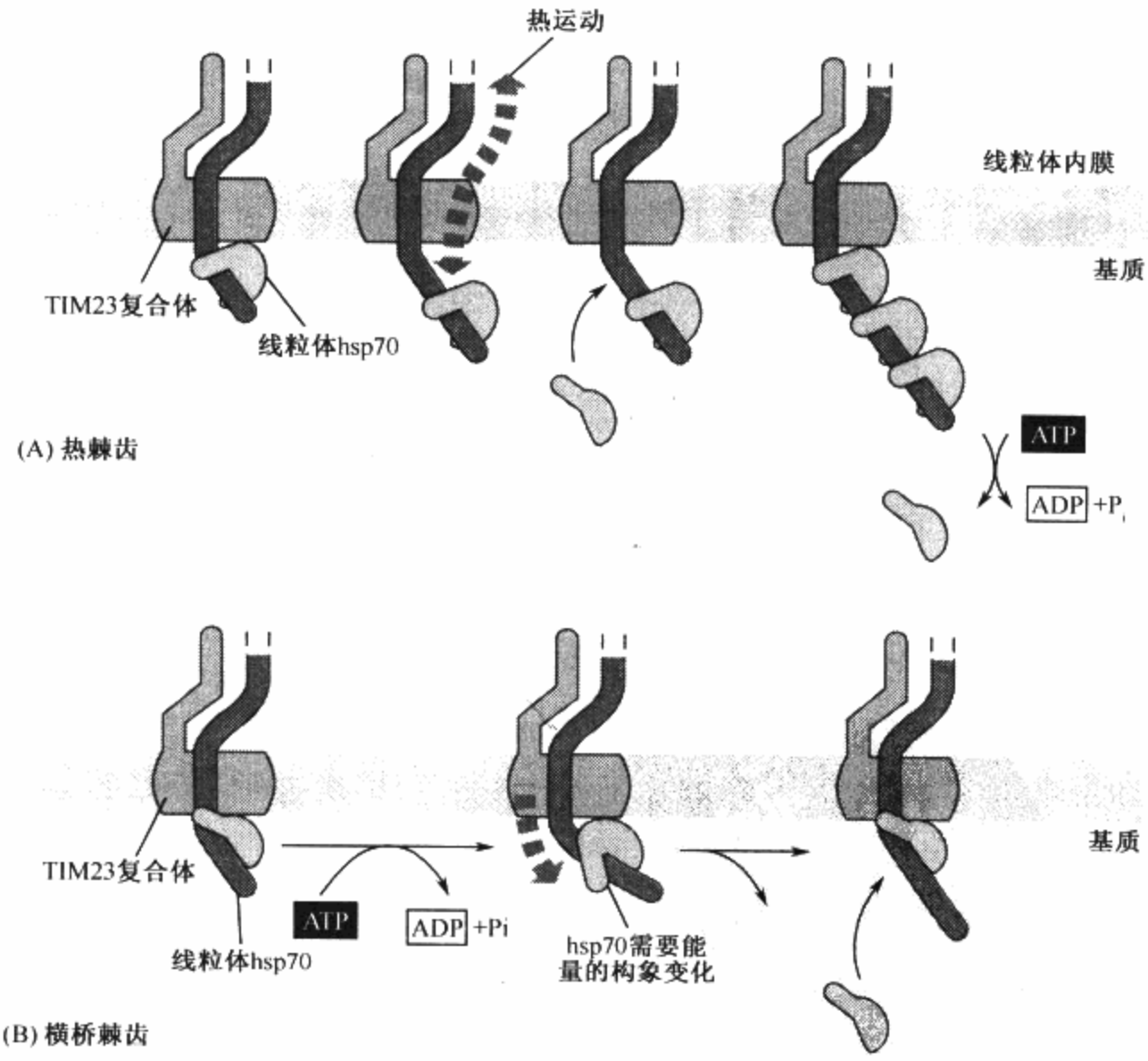


图 12-28　线粒体 hsp70 驱动蛋白质输入的两种可能模型

（A）在热棘齿模型中，正在易位的多肽链由于热运动可以来回滑动，它连续被基质中的 hsp70 结合。（B）在横桥锁止扣模型中，hsp70 的构象改变将多肽链主动拉入基质。在两种模型中，hsp70 都和 TIM23 复合体结合，后者可在易位多肽链刚从复合体中出现进入基质时，将 hsp70 装载在多肽链上

在和线粒体 hsp70 作用之后，多数输入蛋白被传递到另一伴侣蛋白线粒体 hsp60 上。正如第 6 章所述，hsp60 为伸展状态的多肽链提供了一个折叠室，通过 ATP 水解周期结合和释放多肽链以促使其折叠。

转运插入线粒体内膜和进入膜间隙的蛋白质需要 2 个信号序列

最终整合在线粒体内膜上的蛋白质或者在膜间隙中发挥作用的蛋白质最初从细胞质基质转运输入时采用的机制和线粒体基质蛋白质运送过程相同。在一些情况下它们是先被运送到基质中的（图 12-26）。这些蛋白质在其 N 端的信号序列后面紧跟一个疏水性的氨基酸序列。当 N 端序列引导蛋白质输入基质并被基质信号肽酶切除后，疏水序列作为新的 N 端信号序列将利用 OXA 复合体为易位体把蛋白质从插入或运送通过内膜（图 12-24）。OXA 复合体也用于将线粒体编码的蛋白质插入内膜［图 12-29(A)］。细菌细胞膜和叶绿体的类囊体上都发现了极其相似的易位体，它们被认为可以一种相似的机制帮助蛋白质插入膜。

内膜的另一替代途径可避免内膜或膜间隙蛋白都先进入基质［图 12-29(B)］。在这种情况下，TIM23 内膜易位体和紧随 N 端信号序列的疏水序列结合，使后者发挥终止转移序列的功能，防止多肽链进一步通过内膜。在 N 端信号序列被切除后，剩余的部分经过外膜 TOM 复合体被完全拉入膜间隙。不同的蛋白质通过上述两种途径中的一种进入内膜或膜间隙。

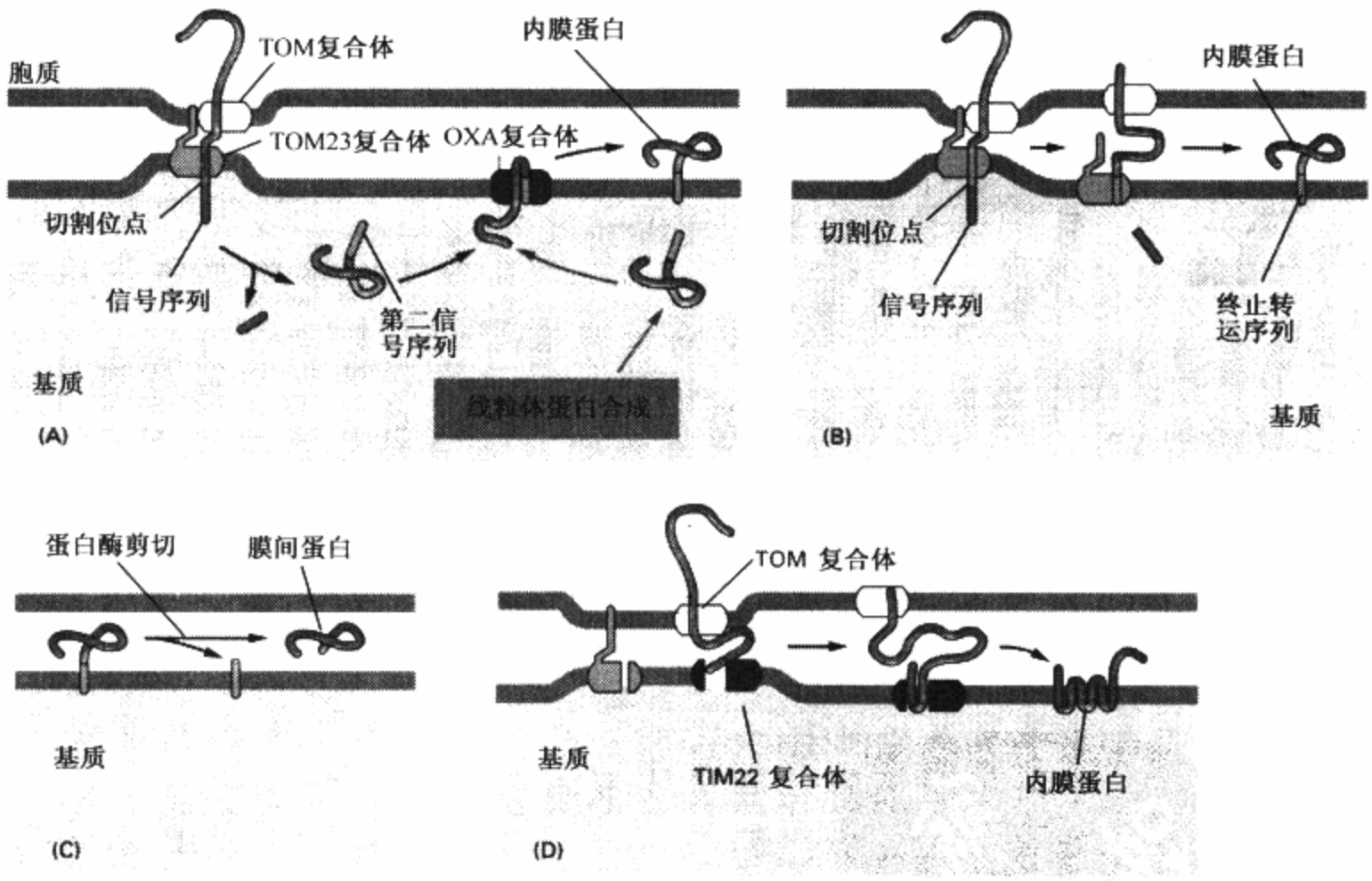


图 12-29　从细胞质基质到插入线粒体内膜或进入膜间隙的蛋白质输入

（A）蛋白质从细胞质基质插入线粒体内膜需要 2 个信号序列和 2 个易位过程。首先前体蛋白输入线粒体基质（图 12-26）。当用于第一次易位的信号序列（红色）被切除后，暴露了处于新的 N 端的疏水信号序列（橙色）。这个信号引导蛋白质进入内膜，其途径可能和线粒体基因组编码蛋白质插入膜所经过的 OXA-依赖型途径相同。（B）在某些情况下，紧随基质定位信号的疏水序列结合内膜上的 TIM23 易位体从而终止易位。剩余的蛋白质经外膜的 TOM 易位体被拉入膜间隙，然后疏水序列被 TIM23 易位体释放至内膜中。（C）一些膜间隙可溶蛋白质在被释放到膜间隙之前可能也需经过（A）、（B）所示的通路，然后其疏水信号序列被第二种活性位点在膜间隙中的信号肽酶切除。（D）用于将代谢物载体蛋白插入线粒体内膜的输入通路利用 TIM 复合体，后者特异性负责多次跨膜蛋白的易位

膜间隙蛋白首先通过疏水信号序列插入内膜，然后被膜间隙的信号肽酶切割释放出成熟的多肽链，以可溶蛋白质的形式进入膜间隙［图 12-29(C)］。这种蛋白质大多作为外周蛋白附着在内膜的外表面上，和跨膜蛋白一起形成蛋白质复合体的亚基。

线粒体是细胞主要的 ATP 合成场所，但是它们也含有许多代谢酶类，如三羧酸循环中的酶。因此线粒体还必须运送除蛋白质之外的小的代谢物通过膜。虽然外膜具有膜孔蛋白，使得小分子可以自由通过，但是内膜不行。大量的小分子是依靠一个代谢特异性的载体蛋白家族的介导通过内膜的。在酵母细胞中，这一家族包括 35 种不同的蛋白质，其中最多的是运输 ADP、ATP 和磷酸的蛋白质。这些内膜载体蛋白是多次跨膜蛋白，它们在 N 端没有可切除的信号序列，而是含有内部转移序列。这些蛋白质通过外膜的 TOM 复合体后再通过 TIM22 复合体插入内膜［图 12-29(D)］。它们整合在内膜上的过程需要 H^+ 的电化学梯度，而不需要线粒体 hsp70 或 ATP。内膜中蛋白质疏水跨膜区域在能学上是有利的，因而可能有助于整合过程的发生。

引导蛋白质插入叶绿体的类囊体膜需要 2 个信号序列

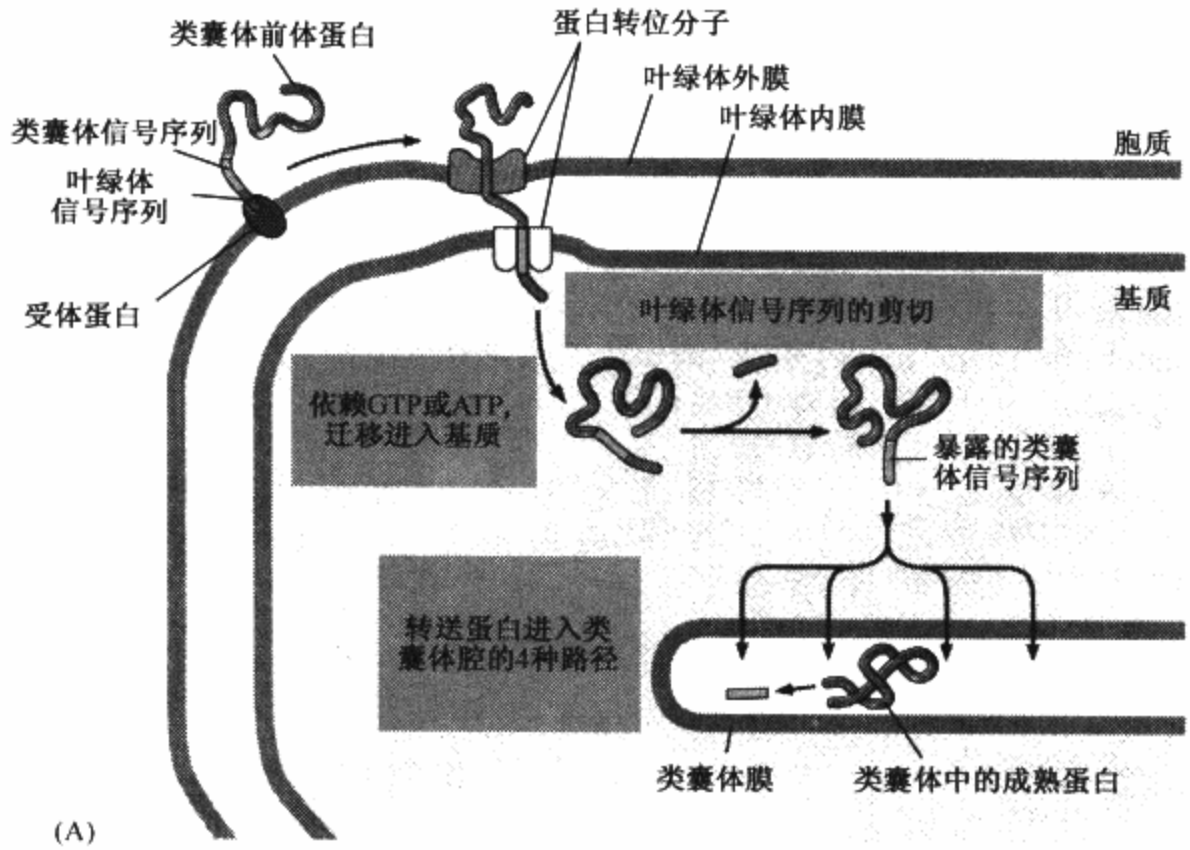
蛋白质输入**叶绿体**和蛋白质输入线粒体在很多方面都很相似。两个过程都发生在翻译后，利用各层膜上独立的易位复合体，通过接触位点，需要能量和两亲性 N 端信号序列，后者在发挥功用后被切除。但是除一些伴侣分子以外，构成两种细胞器易位复合体的蛋白质成分都不相同。此外线粒体利用其内膜两侧的 H^+ 电化学梯度产生转运的动力，叶绿体的 H^+ 电化学梯度存在于类囊体膜两侧而非内膜两侧，因而它们利用 GTP 和 ATP 的水解驱动跨越双层膜的转运。两种细胞器的相似性可能趋同进化的结果，并反映了跨双层膜易位的基本需要。

虽然输入叶绿体需要的信号序列和输入线粒体的信号序列在表面上很相似，但因为在同一植物细胞中同时存在线粒体和叶绿体，因此蛋白质必须选择进入正确的细胞器。例如，在植物体内，如果将细菌酶连接一个线粒体蛋白的 N 端信号序列，则它会被特异性地引导入线粒体；若连接一个叶绿体蛋白的 N 端信号序列，则细菌酶最终定位于叶绿体。因此，不同的信号序列使各细胞器的输入受体可以区分各个待转运的蛋白质。

叶绿体比线粒体多一个膜包被的区域，即**类囊体**。包括光合作用系统蛋白质亚基和 ATP 合酶（讨论见第 14 章）在内的很多叶绿体蛋白质都嵌合在类囊体膜上。和某些线粒体前体蛋白一样，这些蛋白质需经过两个步骤才能从细胞质基质中进入它们的最终目的地。首先这些类囊体膜蛋白从接触位点穿过双层膜进入叶绿体的**基质**，然后再易位插入类囊体膜（或者穿过类囊体膜进入类囊体腔）［图 12-30(A)］。这些蛋白质前体拥有一个紧随 N 端叶绿体信号序列的疏水性类囊体信号序列。N 端信号序列引导蛋白质进入基质后就被基质信号肽酶（线粒体基质信号肽酶的相似物）切除。切除暴露了类囊体信号序列，引发了穿过类囊体膜的转运过程。叶绿体中至少有 4 种途径以供蛋白质穿过或整合在类囊体膜上，4 条途径的基质伴侣分子和能量供给有所不同。

小结

虽然线粒体和叶绿体拥有自身的遗传系统，但它们产生的蛋白质仅占自身蛋白质的一小部分。两种细胞器的大部分蛋白质都是以相似的机制从细胞质基质中输入的。在两种输入过程中，蛋白质都以完全伸展的形式存在。蛋白质通过线粒体被称为接触位点的内外膜黏着处的 TOM 和 TIM 复合体易位进入基质。线粒体易位由 ATP 水解和内膜两侧的 H^+ 电化学梯度驱动发生，而叶绿体易位仅由 GTP 和 ATP 的水解驱动。



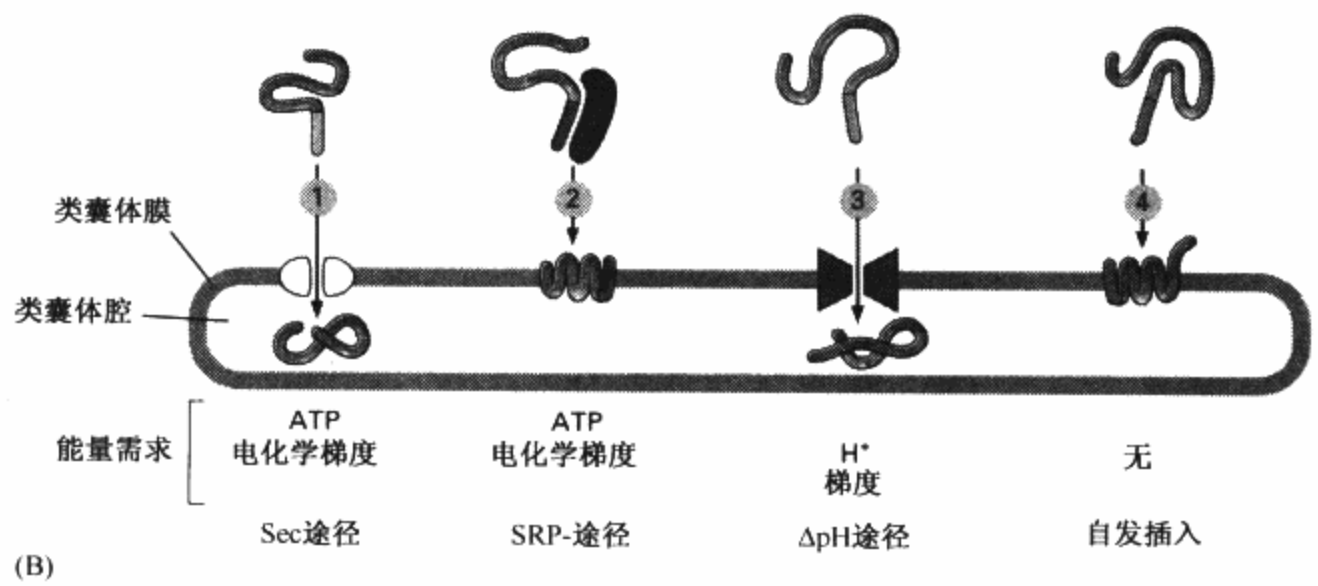


图 12-30　前体蛋白进入叶绿体类囊体的易位

（A）前体蛋白含有一个 N 端叶绿体信号序列（红色），紧随一个类囊体信号序列（橙色）。叶绿体信号序列所引发的叶绿体基质易位也通过膜的接触位点，其机制和线粒体基质易位相似。信号序列随即被切除，暴露类囊体信号序列，引发跨类囊体膜易位过程。（B）进入类囊体腔或插入类囊体膜的易位过程至少可以通过 4 种途径发生：①Sec 途径，因为使用一种 Sec 蛋白的相似体来介导跨细菌质膜的蛋白质易位（下文讨论）而得名。②SRP 状通路，因为使用一种信号识别蛋白的叶绿体相似体而得名。③ΔpH通路，因为由跨类囊体膜的 H^+ 驱动而得名。④自发插入途径，似乎不需要任何用于膜整合的蛋白质易位体

细胞质基质 hsp70 家族的伴侣蛋白使前体蛋白维持在伸展的可易位状态。线粒体或叶绿体基质中的另一组 hsp70 结合刚刚进入的多肽链并将其完全拉进细胞器。只有那些含有特殊信号序列的蛋白质可以被易位进入线粒体或叶绿体。信号序列通常定位于 N 端并在输入后被切除。一些输入蛋白还含有内部转移序列以引导进一步的转运。当最初的信号序列被移除后，这些疏水信号序列便暴露出来引导穿过或插入内膜的第二步转运。在叶绿体中，从类囊体到基质的转运也同样是依赖第二信号，通过几种途径之一完成的。

过氧化物酶体

过氧化物酶体在很多方面都不同于线粒体和叶绿体。其中最明显的一点是它们仅被单层膜包被，并且不含 DNA 或核糖体。但是和线粒体及叶绿体相似的是，过氧化物酶体也通过从细胞质基质的选择性输入获得蛋白质。由于它们没有基因组，所有的蛋白质都依赖于这一转运输入途径。因此作为自我复制、不存在自身基因组的单膜细胞器，过氧化物酶体和内质网更为相似。

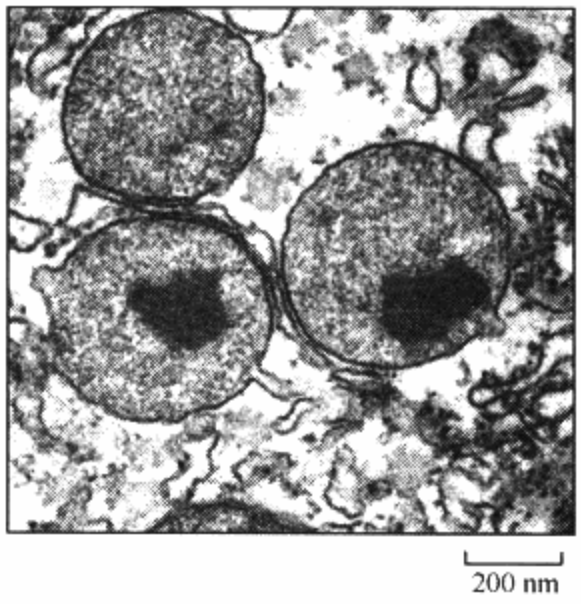


图 12-31　大鼠肝细胞的 3 个过氧化物酶体电镜照片
类结晶的高电子密度包含体由尿酸氧化酶组成

由于其他章节将不再涉及过氧化物酶体，因而在介绍其生物合成过程之前，我们会讨论这一多样细胞器家族的部分功能。所有的真核细胞中都有过氧化物酶体，它们含有浓度极高的氧化酶，如过氧化氢酶和尿酸氧化酶，以至于一些细胞的过氧化物酶体在电子显微镜照片中以类晶体核的形式存在而特别突出（图 12-31）。

和线粒体一样，过氧化物酶体也是细胞主要的耗氧场所。一种假说提出过氧化物酶体是真核细胞最早祖先具有的一种承担所有氧代谢活动的古细胞器遗迹。当光合作用细菌产生的氧开始在大气层中累积时，对大多数细胞都产生了毒害作用。因而过氧化物酶体可能被用来降低细胞内的氧浓度，并利用其化学反应性来进行有用的氧化反应。根据这种观点，线粒体在细胞内的发展必定导致过氧化物酶体大大退化，因为先前发生在过氧化物酶体内的不产生能量的反应可以通过线粒体中的氧化磷酸化偶联于 ATP 而生产。因此现今细胞过氧化物酶体中所发生的氧化反应可能是那些未被线粒体取代的仍具有重要功能的反应。

过氧化物酶体利用氧分子和过氧化氢进行氧化反应

过氧化物酶体通常含有一至多种酶，在氧化反应中这些酶利用分子氧从特异的有机底物（此处设定为 R）中夺走氢原子，产生过氧化氢（H_2O_2）：

$$RH_2 + O_2 \longrightarrow R + H_2O_2$$

过氧化氢酶利用过氧化物酶体中其他酶反应产生的 H_2O_2 氧化多种底物，包括苯酚、甲酸、甲醛和乙醇，反应方程为：$H_2O_2 + R'H_2 \rightarrow R' + 2H_2O$。这种氧化反应是肝脏和肾脏细胞的重要反应，因为其过氧化物酶体负责清除可能进入血液的多种有毒分子。大约 25% 的饮入乙醇通过此种途径被氧化成乙醛。此外，当过量的 H_2O_2 在细胞内累积时，过氧化物酶会通过下述反应将它转变为 H_2O：

$$2H_2O_2 \longrightarrow 2H_2O + O_2$$

过氧化物酶体中氧化反应的主要功能是降解脂肪酸分子。在被称为 β 氧化的代谢过程中，脂肪酸的烃基链通过每次移除 2 个碳原子合成乙酰辅酶 A（acetyl-CoA）的方法被逐步缩短，直至整个脂肪酸转化为乙酰辅酶 A。随后乙酰辅酶 A 从过氧化物酶体中输出进入细胞质基质，参加生物合成反应。在哺乳动物细胞中，β 氧化发生在线粒体和过氧化物酶体中，而在

酵母细胞和植物细胞中，这种必需反应仅发生在过氧化物酶体中。

动物细胞过氧化物酶体的重要生物合成功能是催化合成缩醛磷脂的第一步反应发生。缩醛磷脂是动物体内髓鞘中含量最高的一类磷脂（图 12-32），它的缺乏将引起神经细胞髓鞘形成的严重异常。这也是大多数过氧化物酶体病导致神经疾病的原因之一。

过氧化物酶体通常是一类多样的细胞器，即使是在同一有机个体的各种不同细胞中也可能含有不同的酶类组成。对改变的环境，过氧化物酶体也具有极强的适应能力。例如，酵母细胞一般依赖糖类生长，过氧化物酶体较小。但是使酵母在甲醇中生长时，细胞可以生长出氧化甲醇的大型过氧化物酶体；如果在脂肪酸中生长，则生长出的大型过氧化物酶体将专职于利用 β 氧化将脂肪酸分解为乙酰辅酶 A。

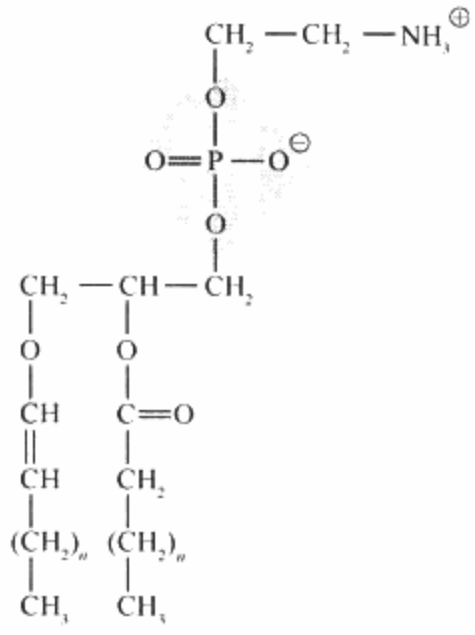


图 12-32　缩醛磷脂的结构

缩醛磷脂在隔绝神经细胞轴突的髓鞘中大量存在。它们组成了髓鞘膜磷脂的 80%～90%。除了一个乙醇胺头部和一个连接在甘油磷酸（所有磷脂的通用骨架分子）上的长链脂肪酸外，缩醛磷脂含有一个特殊的脂肪醇，通过醚键连接在甘油磷酸上（左下角）

过氧化物酶体对于植物也十分重要，其中两种已经进行了广泛的研究。一种是存在于叶片中的过氧化物酶体，它负责催化 CO_2 固定（将 CO_2 固定于碳水化合物中的重要反应）过程的副产物的氧化［图 12-33(A)］。正如第 14 章所述，这一消耗 O_2、释放 CO_2 过程称做光呼吸作用。另一种过氧化物酶体存在于萌发的种子，负责将储存在种子脂类中的脂肪酸转变成幼龄植物生长过程中必需的糖类，因而其地位十分重要。由于从脂类到糖类的转化需要称为乙醛酸循环的一系列反应才能发生，因而这种过氧化物酶体也被称为乙醛酸循环体［图 12-33(B)］。在乙醛酸循环中，过氧化物酶体中脂肪酸降解产生的每 2 分子乙酰辅酶 A 用于丁二酸的合成，新合成的丁二酸离开过氧化物酶体并转化成葡萄糖。动物细胞中没有乙醛酸循环过程，因此动物不能利用脂肪中的脂肪酸产生碳水化合物。

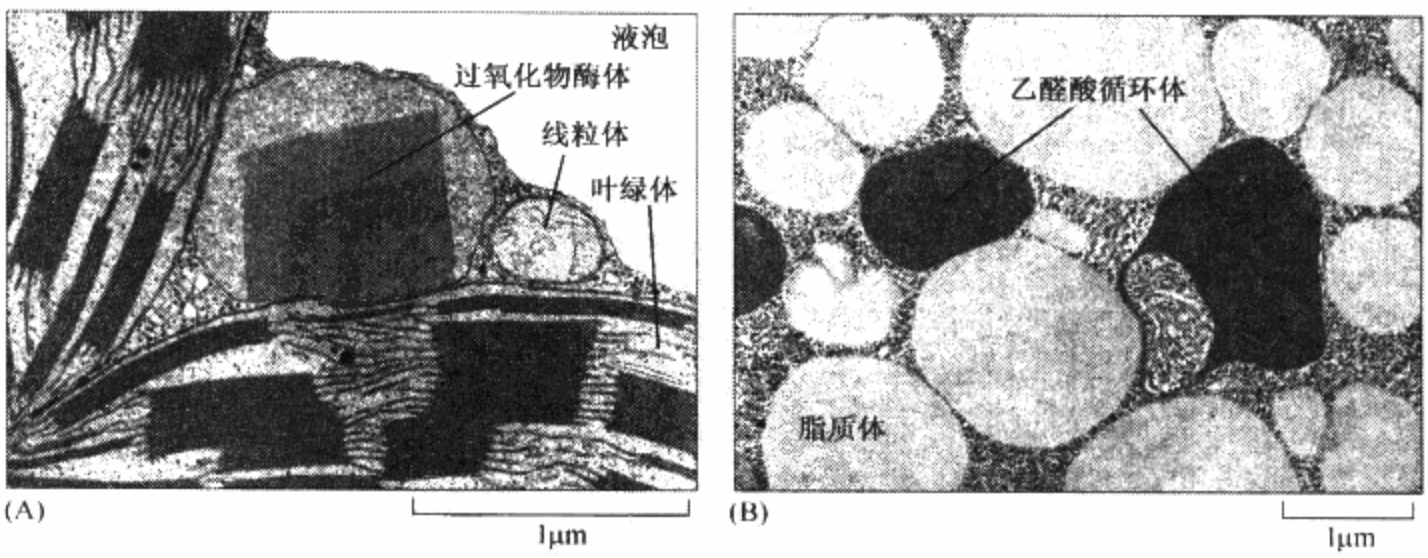


图 12-33　植物细胞中两种过氧化物酶体

（A）烟草叶片叶肉细胞中含有类结晶核的过氧化物酶体。它和叶绿体紧密连接，被认为在光呼吸过程中促进两种细胞器之间的物质交换。（B）萌发后 4 天的番茄种子储存脂肪的子叶细胞中的过氧化物酶体。在此处，过氧化物酶体（乙醛酸循环体）和储存脂肪的脂质体相连，反映了它们在种子萌发过程中的脂肪代谢和糖异生中的中心作用

一个短的信号序列引导蛋白质输入过氧化物酶体

大多数过氧化物酶体蛋白的C端都存在一个特异性的3氨基酸序列，它是过氧化物酶体蛋白质的输入信号（表12-3）。其余少部分过氧化物酶体蛋白在靠近N端的位置上具有信号序列。如果将上述任意一种序列通过实验手段连接在细胞质基质蛋白上，则该蛋白质也可被输入过氧化物酶体。虽然已知输入过程和细胞质基质中识别定为信号的可溶性受体蛋白有关，也和过氧化物酶体胞质面的停靠蛋白有关，但是对整个输入过程的了解依然很少。这一过程至少有23种被称为**peroxin**的特殊蛋白质参与，其能量来源于ATP的水解。由于寡聚蛋白在输入过氧化物酶体时无需展开，说明这一机制和线粒体、叶绿体的输入机制有明显区别。并且至少有一种可溶性输入受体——peroxin Pex5，在整个过程中协同其转运物进入过氧化物酶体，在释放转运物后又返回细胞质基质。蛋白质输入过氧化物酶体的这些特征和蛋白质输入细胞核的过程十分相似。

人类遗传疾病泽韦格综合征（脑-肝-肾综合征）显示了这一输入过程以及过氧化物酶体本身的重要性。这种病症由蛋白质输入过氧化物酶体的缺陷导致严重的过氧化物酶体缺乏引起。患者的细胞仅含有“空载的”过氧化物酶体，脑、肝和肾都严重异常，因而在出生后很快死亡。已经证实，导致这种病症的一种基因突变是编码和蛋白质输入相关的一种过氧化物酶体膜内在蛋白——peroxin Pex2的基因突变。另一种温和的遗传性过氧化物酶体疾病是由于结合N端输入信号的受体缺陷所致。

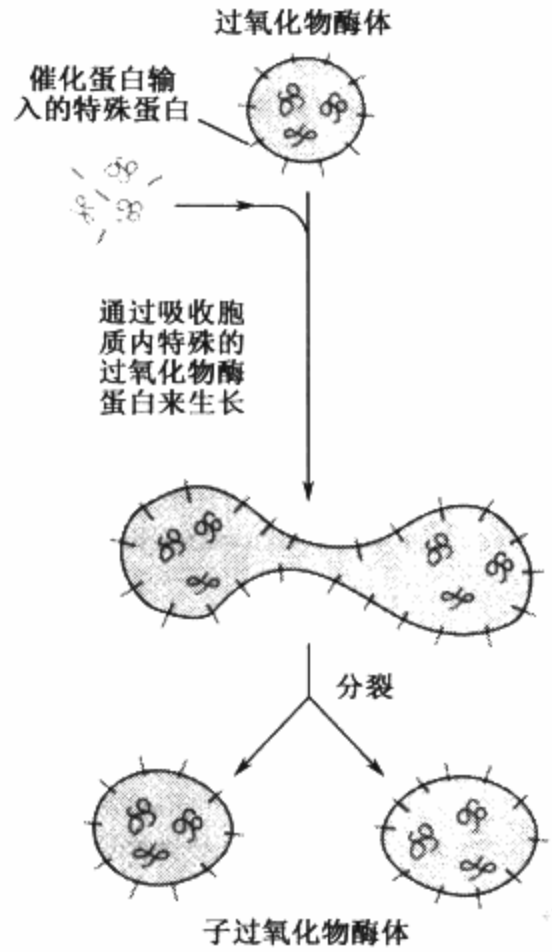


图12-34　新过氧化物酶体产生的模型
过氧化物酶体膜含有输入受体蛋白。包括新的输入受体在内的过氧化物酶体蛋白都是由细胞质基质核糖体合成后再输入细胞器的。组成新的过氧化物酶体膜的脂类也可能经过这一输入途径（将在下文讨论内质网中产生的脂类是如何通过细胞质基质运送到其他细胞器的）。在这个模型中，过氧化物酶体被认为仅通过生长和分裂的过程从前已存在的过氧化物酶体中产生

大多数过氧化物酶体膜蛋白在细胞质基质中产生，然后再插入已存在的过氧化物酶体膜上。因此，新的过氧化物酶体是通过前已存在的个体的生长和分裂产生，这和先前介绍的线粒体及质体相同，也和下文将提到的内质网一样（图12-34）。

小结

过氧化物酶体专门负责利用氧分子进行氧化反应，产生过氧化氢，然后利用后者的氧化性实现必要的目的反应，包括通过过氧化物酶体中过氧化氢酶的催化进行自身氧化还原反应清除过量的过氧化氢。过氧化氢酶体对合成神经细胞髓鞘形成过程所必需的特定磷脂也具有重要作用。和线粒体、质体相似，过氧化物酶体被认为是自主复制的细胞器。但由于它不含有DNA或核糖体，因此必须从细胞质基质中输入蛋白质。很多过氧化物酶体蛋白靠近C端的位置上含有特异性的3氨基酸序列。蛋白质输入基质和线粒体及叶绿体明显不同，寡聚蛋白可在非伸展状态下被转运入过氧化物酶体。

内质网

所有真核细胞都具有**内质网**，一个普通动物细胞中的内质网膜一般占总膜量的一半以上（表 12-2）。内质网是由遍及细胞质基质的分枝状小管和扁平囊构成的网状迷宫结构（图 12-35）。所有小管和囊都互相连接，所以内质网是一层连续的膜包围的细胞内一个独立空间。这个高度蜿蜒卷曲的空间被称为**内质网腔**或内质网池，通常占细胞总体积的 10%以上（表 12-1）。内质网膜将内质网腔和细胞质基质分隔，并协调两个空间分子的选择性转运。

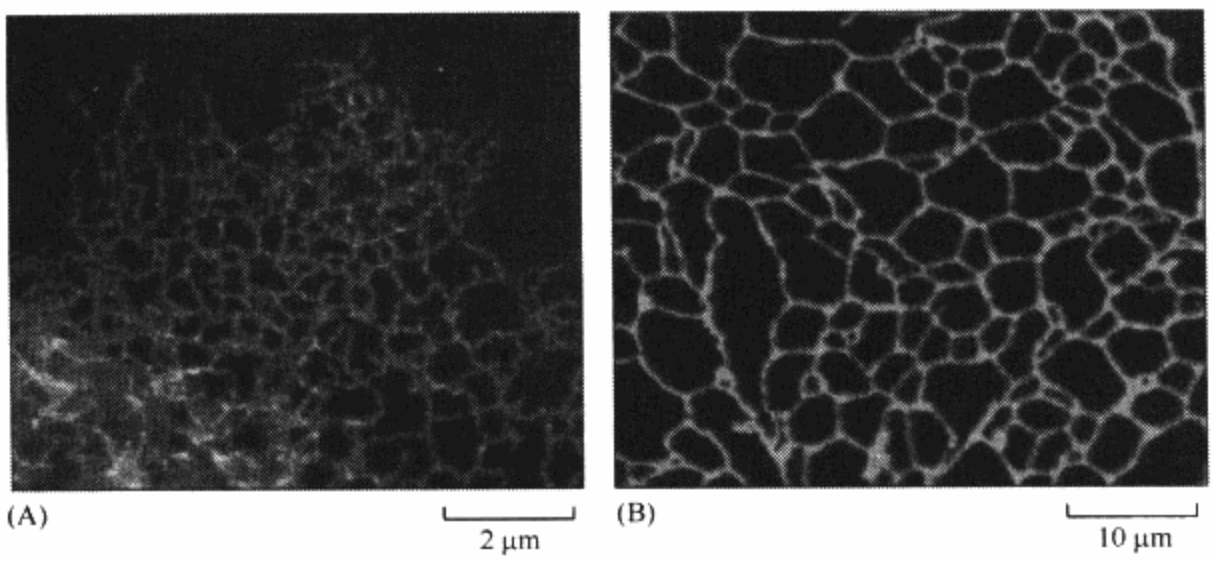


图 12-35　内质网的荧光显微照片

（A）一个培养的哺乳动物细胞中的部分内质网，其蛋白质与抗体结合而使内质网被染色。内质网以网络的形式在整个细胞质基质中伸展，因此细胞内各个区域都和内质网膜的一部分相接近。（B）一个活体植物细胞中的部分内质网，通过遗传工程手段使细胞内质网表达荧光蛋白

内质网的核心功能是脂肪和蛋白质的合成，其膜是大部分细胞器中所有跨膜蛋白及脂类的合成场所，包括内质网自身、高尔基体、溶酶体、内含体、分泌小泡以及细胞膜。内质网生产线粒体和过氧化物酶体膜所需的大部分脂类物质，对其形成起到了主要作用。除此之外，绝大多数分泌至胞外的蛋白质，内质网、高尔基体及溶酶体蛋白最初都先被运送至内质网腔中。

膜旁核糖体定义糙面内质网

内质网从细胞质基质中选择性捕获正在合成的蛋白质。这些蛋白质分为两类：跨膜蛋白，部分易位通过内质网膜并嵌合其中；水溶性蛋白质，完全通过内质网膜进入内质网腔。一些跨膜蛋白始终在内质网上发挥功效，但大部分最终都位于细胞膜或其他细胞器的膜上。水溶性蛋白质将被运送至细胞器腔或被分泌。不管其最终定位在哪里，上述所有蛋白质都被相同的信号序列引导至内质网膜、以相似的机制易位进入内质网。

在哺乳动物细胞中，蛋白质在多肽链合成完全之前就开始输入内质网，即内质网输入是一个**协同翻译**过程。这与蛋白质输入线粒体、叶绿体、核及过氧化物酶体的**翻译后**过程明显不同。由于当蛋白质的一端进入内质网时其余部分的多肽链仍在合成，因而蛋白质从未被释放进入细胞质基质，在到达内质网膜上的易位体之前也不存在折叠的危险。所以，和线粒体、叶绿体的蛋白质翻译后输入过程相比，内质网易位不需要伴侣蛋白来维持蛋白质的伸展状态。合成蛋白质的核糖体直接附着在内质网膜上。这些膜旁核糖体包被在内质网表面，产生**糙面**

内质网［图 12-36(A)］。

细胞质基质中两种空间位置不同的核糖体。膜旁核糖体附着在内质网膜的胞质面，负责合成的蛋白质在合成过程中同时易位进入内质网。游离核糖体不与任何膜结合，合成核基因组编码的其他蛋白质。膜旁核糖体和游离核糖体在结构和功能上相同，它们的区别仅是合成不同的蛋白质。当一个核糖体恰好合成一个带有内质网信号序列的蛋白质时，信号序列将核糖体引导至内质网膜上。

因为很多核糖体可以同时结合在一条 mRNA 链上，所以它们经常形成**多核糖体**，并被正在延伸的多肽链的信号序列引导至内质网膜上［图 12-36(B)］。结合在这种 mRNA 分子上的核糖体在完成 mRNA 的 3′端翻译后可以返回细胞质基质中。但是 mRNA 分子仍然被其他的核糖体黏附在内质网膜上，这些核糖体都是被易位体短暂地固定于膜上的。相反，如果一个 mRNA 分子编码的蛋白质没有内质网信号序列，则形成的多核糖体始终处于游离状态，并直接释放蛋白质于细胞质基质中。因此，只有那些编码含内质网信号序列蛋白质的 mRNA 分子结合粗面内质网膜；而编码其他蛋白质的 mRNA 分子游离于细胞质基质中。而核糖体亚基个体则在两种空间位置隔离的 mRNA 分子间随机运动（图 12-37）。

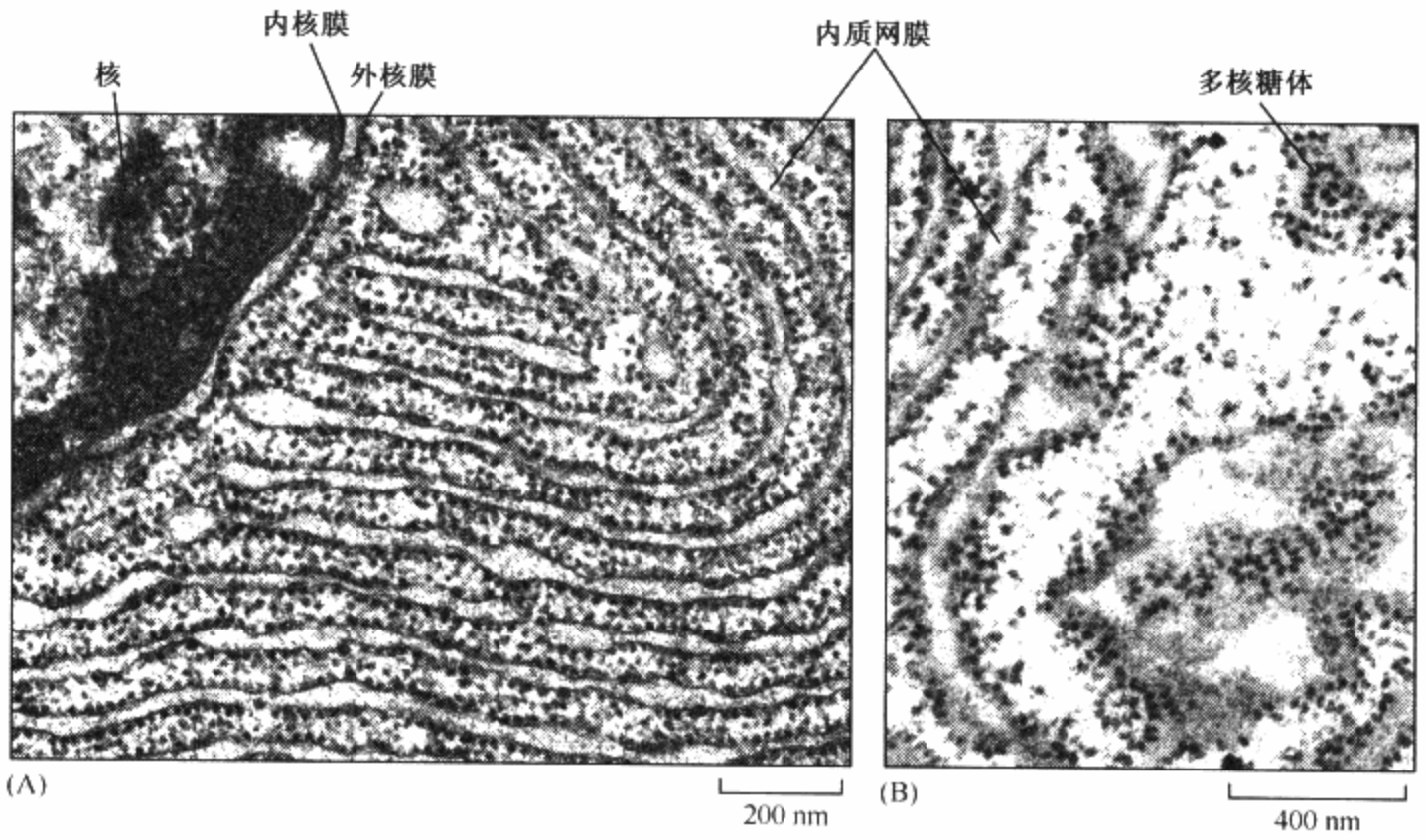


图 12-36　粗面内质网

（A）一个胰腺分泌细胞的粗面内质网电镜照片。胰腺分泌细胞每日产生并分泌大量消化酶，在电镜照片中，其细胞质基质充满膜上附着核糖体的密集排列的内质网。在图的左上是细胞核及其核被膜的一部分；注意外核膜和内质网膜相连接，并且其上也附着有核糖体。（B）附着在内质网上的多核糖体的薄切片电镜照片。截面显示在一些区域，平行于膜的内质网被切断，显示出多核糖体的玫瑰花环形态

光面内质网在某些特殊细胞中大量存在

缺乏膜旁核糖体的内质网成为**光面内质网**。在大多数细胞中，这种内质网通常极少并且半光滑半粗糙。光面内质网含有内质网出口位点，以供装载了新合成的蛋白质和脂类的运输小泡出芽运送到高尔基体，因此它们有时又称为过渡内质网。但是在某些特殊细胞中，光面内质网大量存在并且具有额外的功能，尤其是在那些专门负责脂类代谢的细胞中特别突出。例如，利用胆固醇合成类固醇激素的细胞具有扩大的光面内质网来储存生产胆固醇和加工它

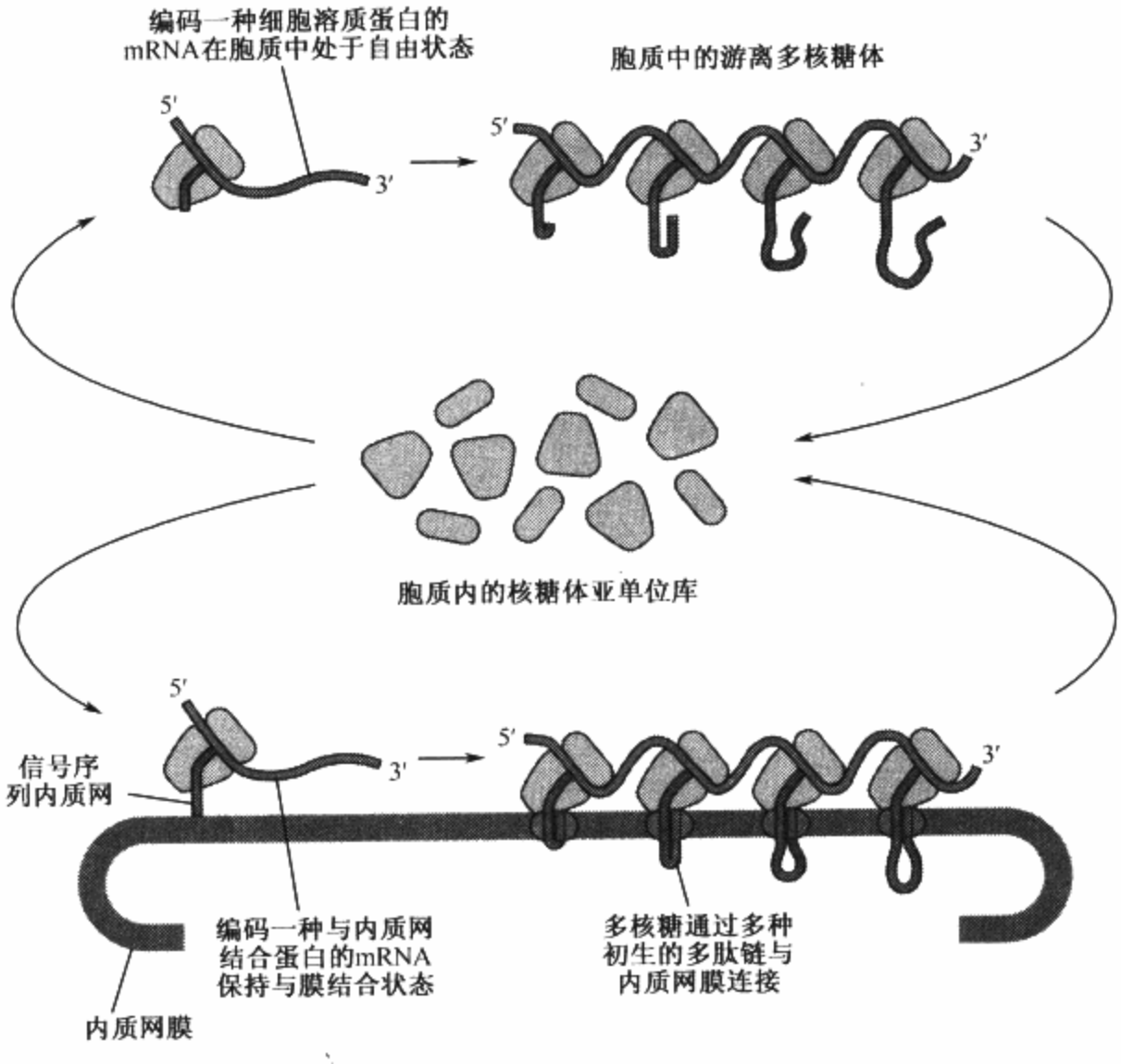


图 12-37　游离核糖体和膜结合核糖体

核糖体库用于合成停留在细胞质基质中的蛋白质和转运进入内质网的蛋白质。新合成的多肽链上的内质网信号序列引导合成它的核糖体至内质网膜。当多个核糖体循环通过 mRNA 分子时，后者始终作为多核糖体的一部分结合在内质网上。在每一次蛋白质合成完成后，核糖体亚基都会被释放然后重新回到细胞质基质中的核糖体库中

成为激素所必需的酶［图 12-38(A)］。

肝脏中的主要细胞肝细胞是另一种含有大量光面内质网的细胞，它是产生脂蛋白颗粒的主要场所，脂蛋白能将脂类从血液中运送至机体的其他部分。合成脂蛋白脂类部分的酶定位在光面内质网的膜上，除此之外，脂溶性药物和众多代谢有害产物的解毒反应的催化酶也在光面内质网中。解毒反应中研究得最多的是细胞色素 P450 家族酶系，它催化一系列反应使得

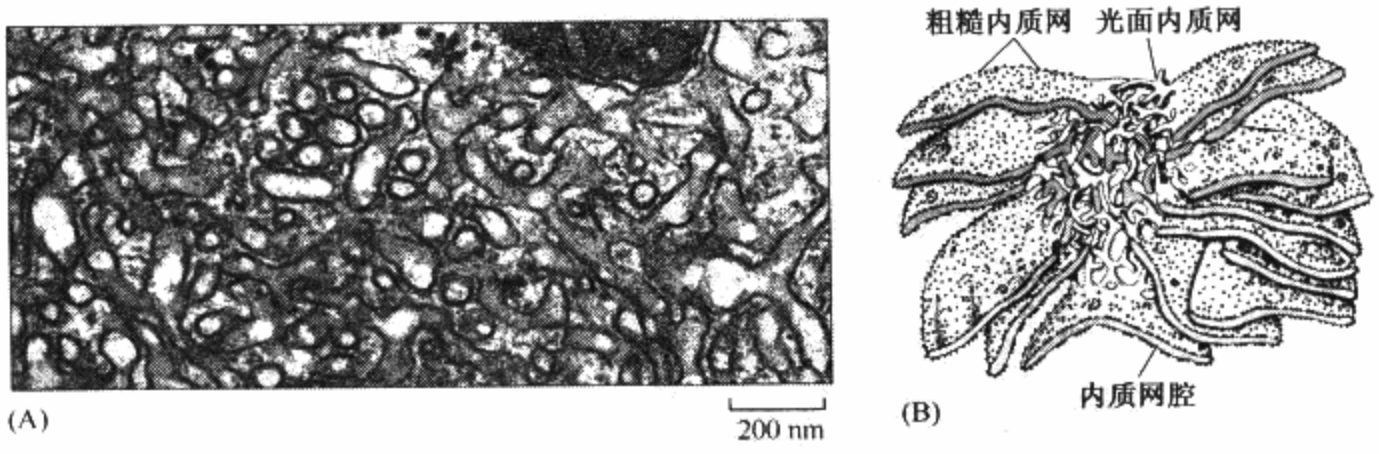


图 12-38　光面内质网

（A）类固醇激素分泌细胞中的大量光面内质网。这张电镜照片显示人睾丸中分泌睾酮的 Leydig 细胞。（B）肝脏细胞光面内质网和粗面内质网区域的三维结构重建。粗面内质网形成多摞有方向的扁平池，每个池管腔为 20～30nm 宽。光面内质网和这些池相连形成直径为 30～60nm 的管道网络结构

水溶性药物或代谢物尽可能地溶于水中离开细胞以尿的形式排泄出去，否则这些有害物质将在细胞膜系统上累积。由于粗面内质网不能容纳足够的这种酶以及其他必要的酶类，通常肝细胞中内膜系统主要由光面内质网构成［图 12-38(B) 和表 12-2］。

当大量的特定化合物，例如，药物苯巴比妥进入机体循环时，肝脏中会大量合成解毒反应的酶，光面内质网的表面积在几天之内就会成倍增加。一旦药物消失，多余的光面内质网膜会特异且迅速地被一个称为自吞噬作用的溶酶体依赖过程清除（讨论见第 13 章）。但是目前尚不知这一巨大的变化过程是如何被调节的。

大多数真核细胞光面内质网的另一功能是把 Ca^{2+} 和细胞质基质分隔开。正如第 15 章所述，很多对细胞外信号的迅速反应都离不开内质网的 Ca^{2+} 释放以及紧随的 Ca^{2+} 从细胞质基质中重摄取的过程。内质网中具有高浓度的钙结合蛋白，这促进了内质网腔的 Ca^{2+} 储存。在某些甚至大部分细胞类型中，内质网的特定区域专门负责 Ca^{2+} 的储存。例如，肌细胞具有大量专门化的光面内质网，称为肌质网，它利用 Ca^{2+}-ATP 酶将 Ca^{2+} 从细胞质基质中泵入内质网腔，使 Ca^{2+} 和细胞质基质隔离。在每一次肌收缩过程中，肌质网的 Ca^{2+} 释放和重摄取分别引起肌原纤维的收缩和舒张（讨论见第 16 章）。

下面我们将讨论内质网的两个主要功能：合成修饰蛋白质以及合成脂类。

糙面内质网和光面内质网可以通过离心分离

为了研究内质网的生化功能，分离内质网膜是十分必要的。由于内质网和细胞质基质中的其他成分杂乱地掺杂在一起，因此这看起来似乎是不可能完成的任务。所幸当组织或细胞经过匀浆破碎后，内质网会断裂为碎片并重新结合为很多封闭的小囊泡（直径 100～200nm），称为**微粒体**。这些微粒体相对容易纯化。来源于粗面内质网的微粒体布满核糖体，称为粗糙微粒体。核糖体一般总是在外表面，因此微粒体的内表面在生化特征上和内质网腔面相同（图 12-39）。因为它们可以容易地以功能型的形式纯化出来，粗糙微粒体对于研究粗面内质网中进行的生化过程极为有用。对于生物化学家来说，它们是粗面内质网的微缩模型，可以进行蛋白质的合成、糖基化，Ca^{2+} 的摄取以及脂类的合成。

在匀浆液中还可以找到很多和粗糙微粒体大小相似但缺乏附着的核糖体的囊泡。这些光滑微粒体部分来源于光面内质网，部分来源于细胞膜、高尔基体、胞内体以及线粒体的片段（成分比例取决于组织）。因此，和粗面内质网碎片产生的粗糙微粒体相比，光滑微粒体的来

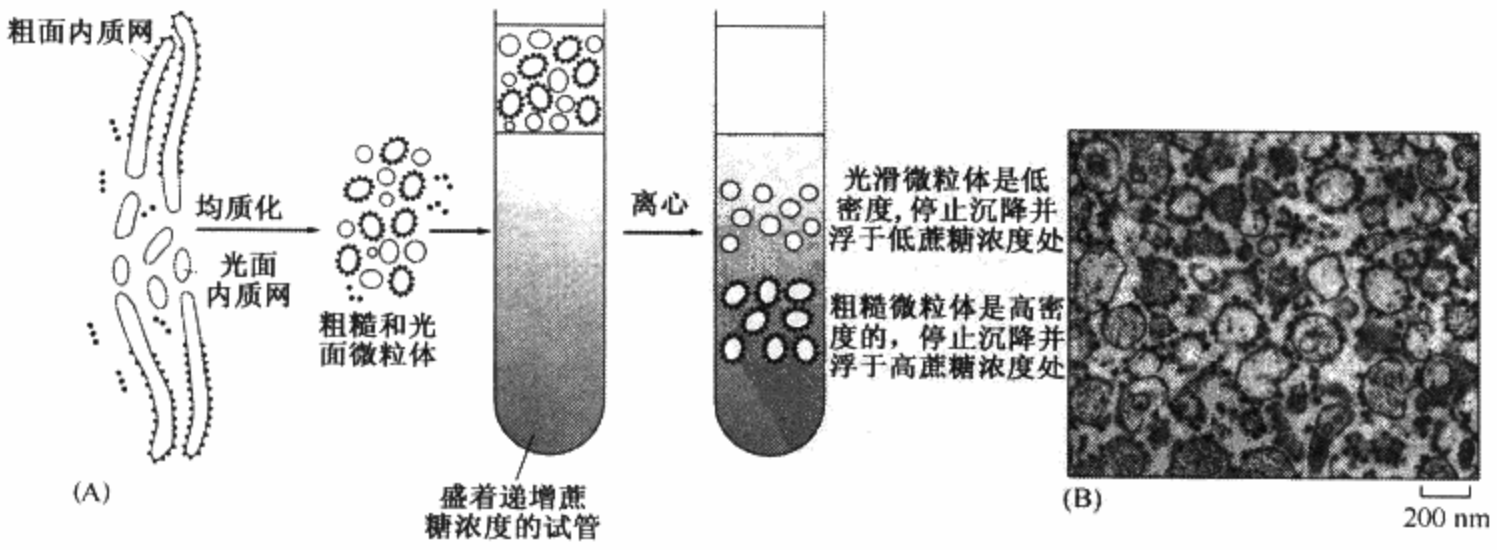


图 12-39　分离从内质网纯化出的粗糙微粒体和光滑微粒体

（A）当用梯度蔗糖溶液将两种微体沉淀至平衡状态时，可以根据其不同的密度将它们分开。（B）纯化的粗面内质网碎片薄切片的电镜照片显示大量核糖体附着的囊泡

源难以判定。但是肝脏的微粒体是一个例外，因为肝细胞的光面内质网数量出奇的多，肝脏匀浆液中的大多数光滑微粒体来源于光面内质网。

附着的核糖体使得粗糙微粒体的密度大于光滑微粒体［图 12-39(B)］，因此通过均衡离心可以将它们彼此分离。当比较分离的肝脏粗糙微粒体和光滑微粒体的酶活性、多肽链成分等其他性质时，它们十分相似但不尽相同：显然由于内质网膜是一个连续、流动的膜，其膜成分可以在粗糙和光滑区域间自由传播。但是粗糙微粒体比光滑微粒体多含有 20 种蛋白质，说明存在一个针对一类内质网膜蛋白的分离机制。这类蛋白质中的一部分帮助核糖体结合在粗面内质网上，其余可能使粗面内质网部分成为扁平状［图 12-38(B)］。尚不清楚这类蛋白质是通过在类脂双层膜中形成二维的集合体还是通过和内质网膜两面的结合蛋白网络相互作用被限定在粗面内质网上的。

信号序列是在输入糙面内质网的蛋白质中首次发现的

信号序列（以及蛋白质分选的信号序列机制）是在 20 世纪 70 年代早期在分泌蛋白中发现的。这些分泌蛋白首先易位穿过内质网膜，最终被细胞释放至胞外。在关键实验中，编码分泌蛋白的 mRNA 被体外核糖体翻译。当这个无细胞系统中不加入微粒体时，合成的蛋白质比正常的分泌蛋白略大一点，多出的长度即是 N 端前导肽。当加入来自粗面内质网的微粒体时，系统合成了正确大小的蛋白质。信号假说揭示了这一实验结果：前导肽作为**内质网信号序列**引导分泌蛋白至内质网膜，随后在多肽链翻译完成之前被内质网膜中的信号肽酶切除（图 12-40）。

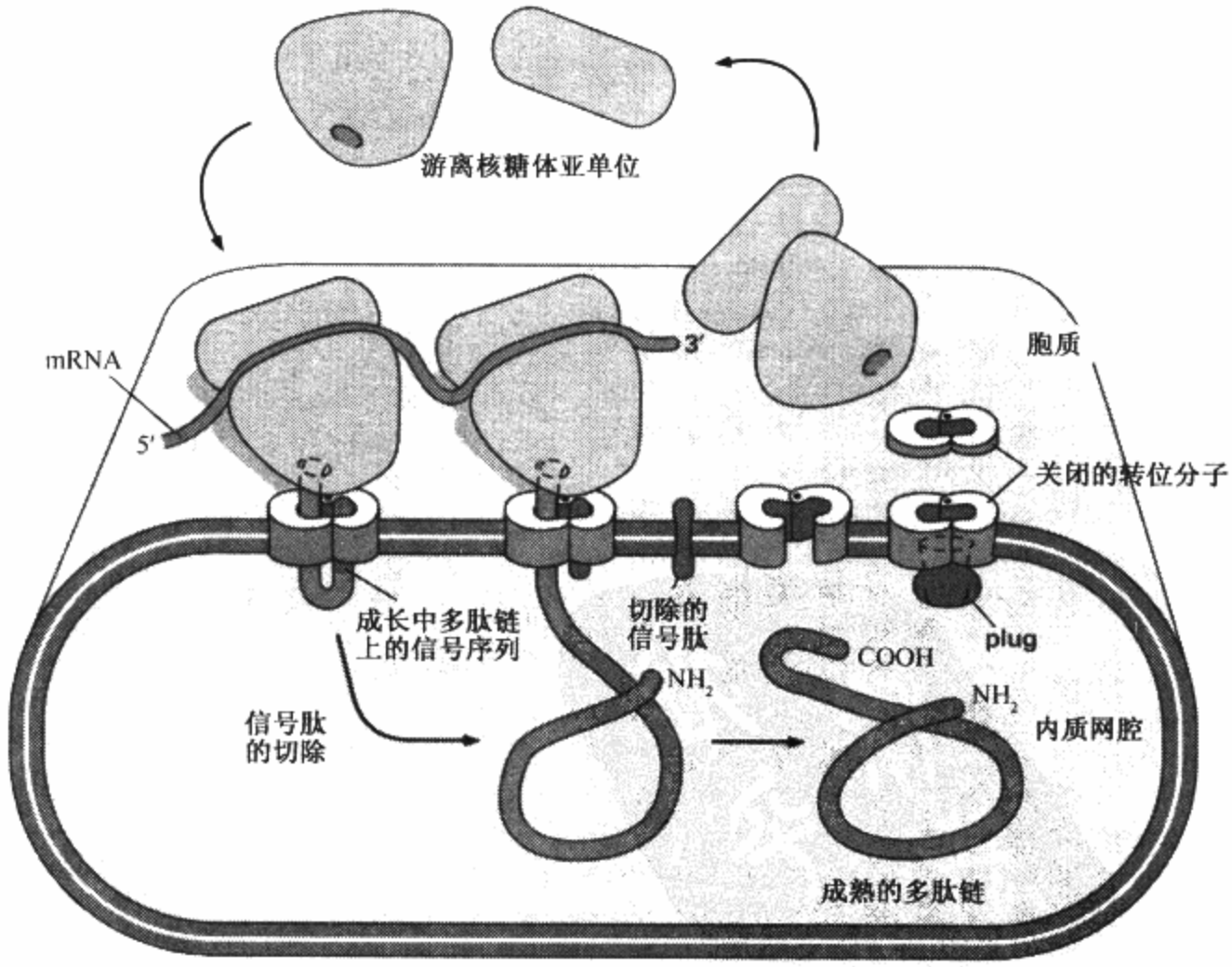


图 12-40　信号假说

按最初假设绘制的跨内质网膜蛋白质易位简易模式。当内质网信号序列从核糖体中露出后，会引导核糖体至内质网膜上的易位体，易位体在膜上形成一个孔道使多肽链从此处通过。在易位过程中，信号序列被信号肽酶移除，成熟的蛋白质在合成后即刻便被释放至内质网腔。我们现在知道这个假设大体上是正确的，但是还有一些图中未显示的附加成分

根据信号假说，分泌蛋白在体外合成过程中应进入微粒体腔。这可利用蛋白酶的作用证实：当体系中不存在微粒体时，加入的蛋白酶可以将新合成的蛋白质降解；而微粒体存在的条件下，由于微粒体膜的保护，相同的蛋白质可以保持完整。但是当不含内质网信号序列的蛋白质在细胞外体系中合成后，它们不能输入微粒体因而会在蛋白酶作用的情况下降解。

信号假说已经得到了遗传学实验和生化实验的完全证实，并且它对植物细胞、动物细胞以及跨细菌细胞膜、线粒体膜、叶绿体膜和过氧化物酶体膜的蛋白质易位都是普适的。N 端内质网信号序列不仅介导可溶性分泌蛋白，并且介导其他所有产生于粗面内质网上核糖体的蛋白质前体易位，包括膜蛋白。这些序列的信号功能是通过重组 DNA 技术被直接证实的：将内质网信号序列连接在本来不含有它们的蛋白质上，结果是融合蛋白被引导进入内质网。

无细胞系统中，蛋白质输入微粒体。这为鉴定、纯化和研究各种负责内质网输入过程分子机制中的各种成分提供了强有力的分析手段。

一种信号识别蛋白(SRP)引导内质网信号蛋白结合糙面内质网膜上特异性受体

内质网信号序列被至少两种成分引导至内质网膜：在内质网膜和细胞质基质之间循环的**信号识别蛋白**（SRP），它可以结合信号序列；以及内质网膜上的**信号识别蛋白受体**。信号识别蛋白是一个由 6 条不同的多肽链结合一条小 RNA 单链分子组成的复合体［图 12-41(A)］。SRP 的同系物及其受体在所有被研究过的有机体中都有发现，这说明这种蛋白质定位机制在进化过程中很早就出现了，并且保留至今。

内质网信号序列在氨基酸序列上有很大差异，但是每一种都在其中心位置含有 8 个以上非极性氨基酸（表 12-3）。SRP 如何特异性地结合这么多不同的序列？答案来自 SRP 蛋白的结晶结构，它显示信号序列的结合位点是一个连接甲硫氨酸的大型疏水口袋［图 12-41(B)］。由于甲硫氨酸的残基不分支且易弯曲，因而这个具有相当大可塑性的口袋可以容纳不同序列和形态的疏水信号序列。

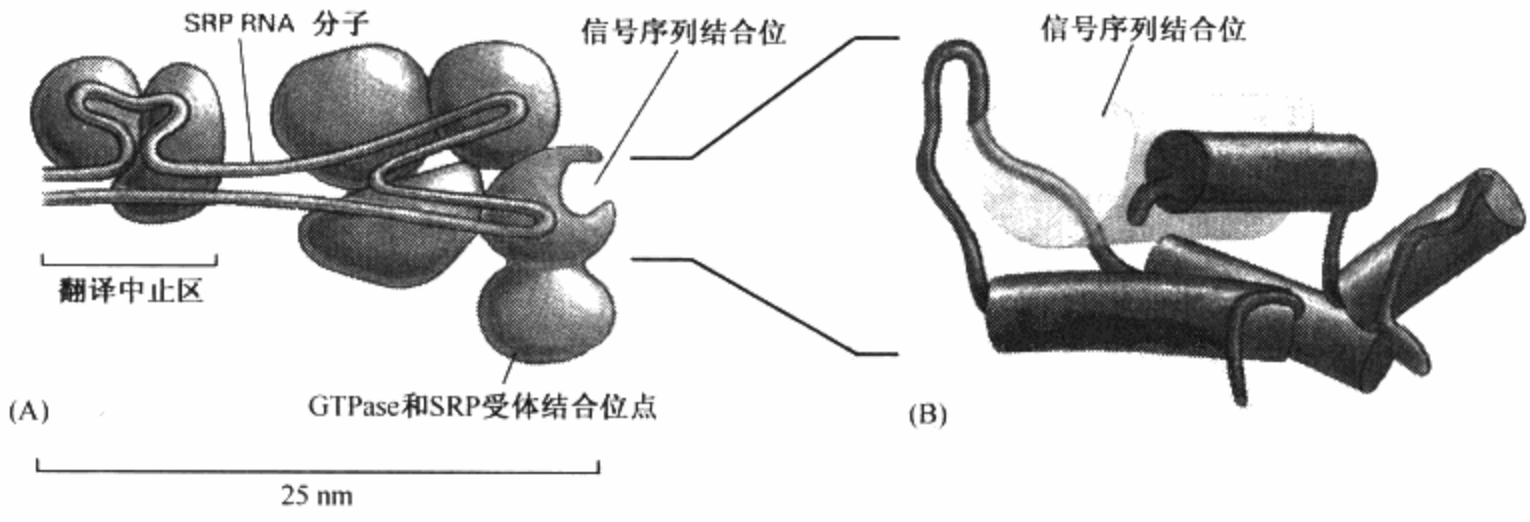


图 12-41　信号识别蛋白（SRP）

（A）哺乳动物的 SRP 是一种包含 6 个蛋白亚基和 1 个 RNA 分子（SRP RNA）的复合体。SRP 的一端结合在正在生长的多肽链的信号序列上，而另一端结合在核糖体上中止翻译过程。信号识别蛋白中的 RNA 可能介导了 SRP 和核糖体 RNA 的反应。（B）细菌 SRP 亚基的信号序列结合功能结构域的结晶结构。这一功能结构域具有一个大型的、暴露的结合袋，由疏水性氨基酸构成，其中主要是甲硫氨酸。袋用灰色阴影凸现出来以强调其位置。甲硫氨酸灵活的侧链非常适于构建适应其他不同蛋白质的疏水性结合位点。例如，钙调蛋白（如第 15 章中所述）可结合很多靶蛋白，并且和 SRP 一样具有很多甲硫氨酸聚集域以结合固定不同形状的目标（图 15-40）

当多肽链一出核糖体，SRP 就会结合在内质网信号序列上。这会引起蛋白质合成的暂停，暂停可能使得核糖体在多肽链完成合成之前有足够的时间结合到内质网膜上，这样就保证蛋白质不被释放到细胞质基质中。这一安全的结构可能对分泌水解酶和溶酶体水解酶尤为重要，因为一旦它们被释放到细胞质基质中，将引起严重的破坏；此外，分泌大量水解酶的细胞还有其他防护措施：在它们的细胞质基质中存在高浓度的水解酶抑制体。

一旦 SRP-核糖体复合物形成，SRP 就会和受体结合。SRP 受体是粗面内质网膜上的膜内在蛋白，仅部分暴露于胞质面。SRP 及其受体的相互作用将 SRP-核糖体复合体带至蛋白质易位体，随后 SRP 和 SRP 受体被释放，仍在合成生长的多肽链穿过内质网膜（图 12-42）。

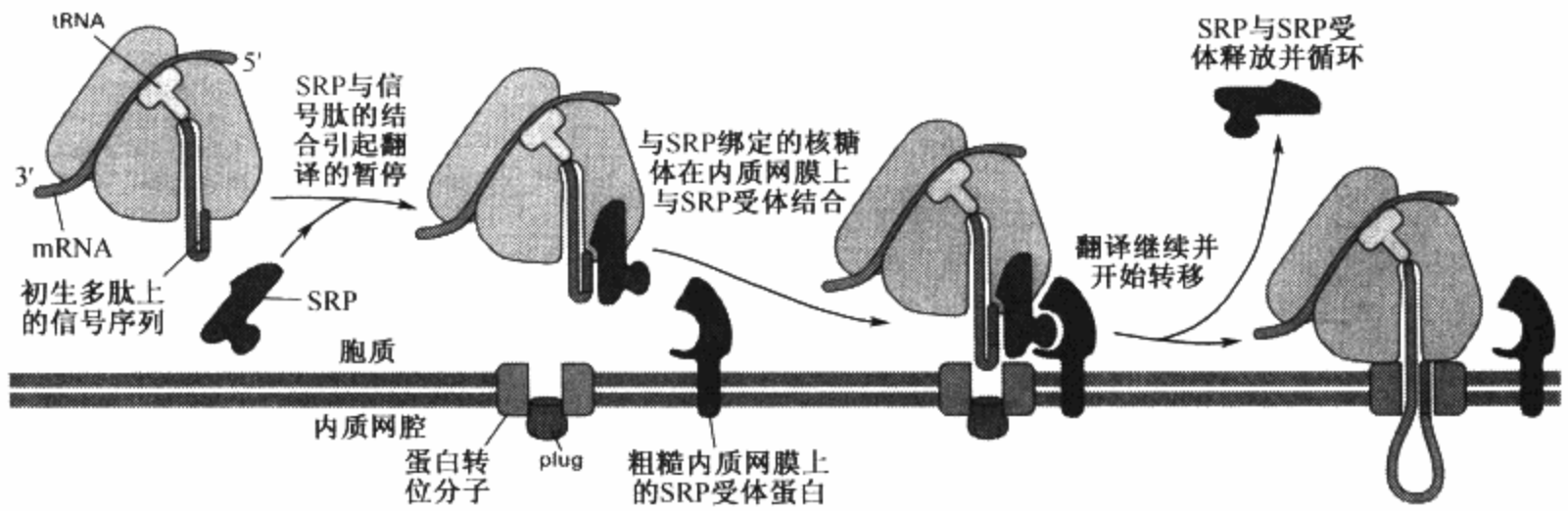


图 12-42　内质网信号序列和 SRP 如何引导核糖体至内质网膜

内质网及其受体共同作用。SRP 同时结合核糖体和暴露出的内质网信号序列，由此引发翻译的中止。内质网上的 SRP 受体由两种不同的多肽链组成，它结合 SRP-核糖体复合体并引导其至易位体处。后续的反应至今了解不多，SRP 和 SRP 受体从核糖体上释放下来，留下结合在内质网易位体上的核糖体。然后易位体使多肽链插入并通过内质网类脂双层膜。因为 SRP 的一条链和 SRP 受体的两条链都具有 GTP 结合功能结构域，因此伴随着 GTP 结合和水解循环过程（讨论见第 15 章）发生的构象转变被认为确保了 SRP 的释放仅发生在核糖体正确的结合了内质网膜上的易位体之后。易位体一直是关闭的（内质网腔面塞原理显示），直到内质网结合时才打开，因此内质网始终存在透过屏障

多肽链通过易位体的含水孔道

多肽链在跨内质网膜的过程中究竟是直接和类脂双层膜接触还是通过蛋白质易位体孔道是一个争论了很久的问题。最终这个争论以蛋白质易位体的纯化而宣告结束。易位体在膜上形成了一个充满水的孔道，多肽链通过这个孔道穿过膜。称为 **Sec61 复合体**的易位体由 3～4 个蛋白质复合体组成，每一个复合体都由 3 个跨膜蛋白组装成一个环形结构。

当核糖体结合其上时，易位体的中心孔道和核糖体大亚基的通道连通，使得正在伸长的多肽链从核糖体中进入易位体（图 12-43）。核糖体和结合的易位体之间紧紧密封不允许任何分子逃离内质网，好像核糖体中的空间和内质网腔中的空间是自然连续的（图 12-44）。但是易位体上的孔道并非始终打开；否则 Ca^{2+} 可能在核糖体分离时泄漏出来。因此普遍认为内质网腔蛋白可能充当了塞子；或者易位体自身在没有核糖体结合时可以发生重排，自动关闭孔道。这样，易位体孔道就成为一个动态结构，仅当正在合成多肽链的核糖体和内质网膜结合时，才开放片刻。

正在生长的多肽链中的信号序列被认为可以引起孔道的开放：在信号序列从 SRP 中被释放出来并且多肽链已经延伸足够的长度之后，信号序列和孔道内的一个特殊位点结合，引发孔道开放。因此，一个内质网信号序列经过了两次识别：首先被细胞质基质中的 SRP 识别，再被内质网蛋白易位体中的结合位点识别。这可能有助于确保只有那些适宜的蛋白质进入内质网腔。

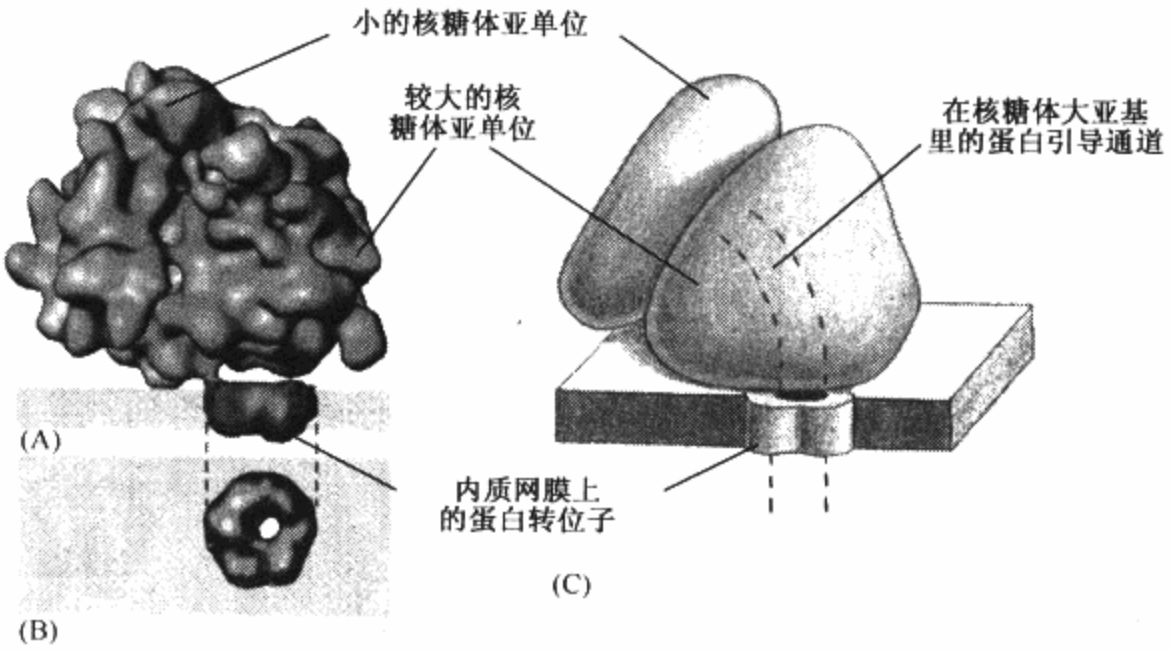


图 12-43　结合在 Sec61 蛋白易位体上的核糖体

（A）复合体侧向电镜照片的重建。（B）易位体俯视图（向内质网膜方向俯视）。（C）结合于易位体的膜结合核糖体示意图。易位体中央的孔道和核糖体大亚基上的通道对接，通过这条通道正在生长的多肽链离开核糖体［见图 6-68(C)］

（A 和 B 来自 R. Beckmann et al., *Science* 278：2123—2126，1997.）

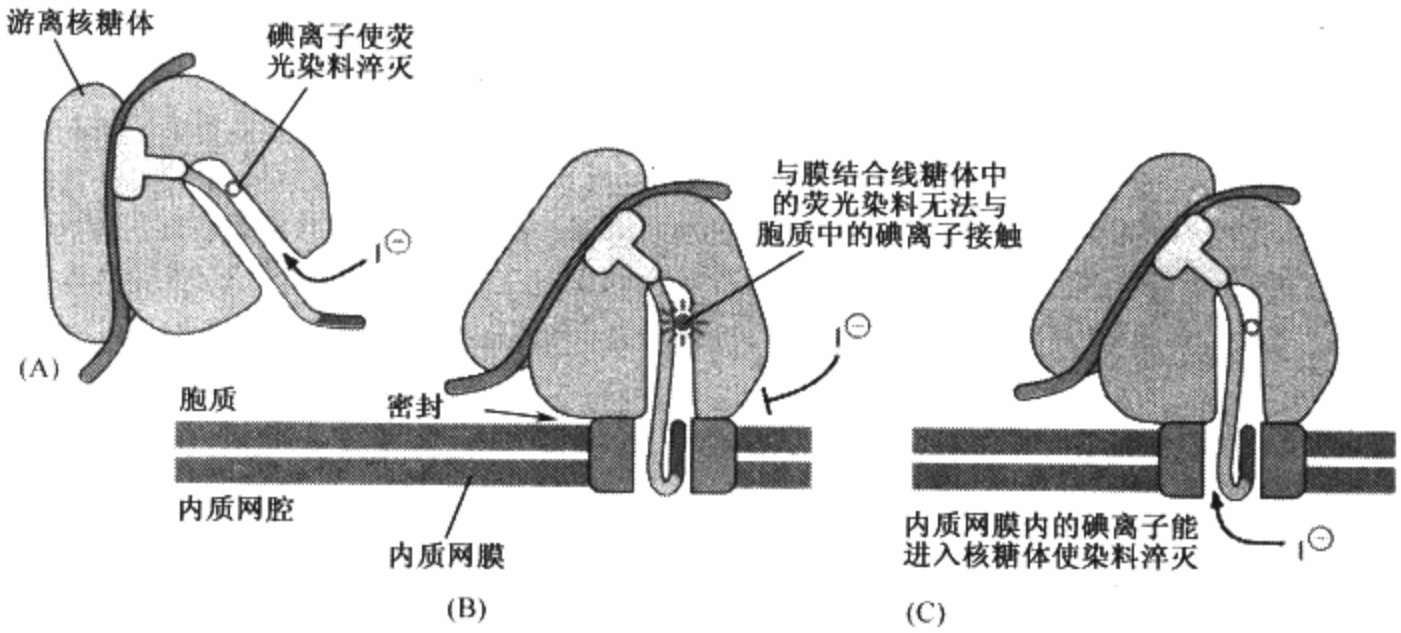


图 12-44　连续水孔道连接内质网腔和核糖体内部的证据

在这一实验中，荧光染料附着延伸多肽链的处于核糖体内部的一段上。（A）在游离核糖体中，染料易于接触细胞质基质溶液中的碘离子。当这种离子和染料结合时，可使荧光染料淬灭。（B）相比而言，对于可以和内质网形成紧密连接的膜结合核糖体，就可以防止碘离子和染料相接触。（C）当向内质网腔加入碘离子时，它们可以通过易位体一路进入到核糖体通道中淬灭膜结合核糖体内部的染料

内质网膜易位并不总要求多肽链同时进行延长

正如已经知道的，线粒体、叶绿体和过氧化物酶体的蛋白质以为都发生在转录后，即蛋白质已经被合成并被释放至细胞质基质之后。而内质网易位通常发生在转录过程中。这就解释了为什么结合与内质网的核糖体一般不结合在其他细胞器上。

但是仍有一些蛋白质是在其完全合成后输入内质网的，说明内质网易位并不总是要求正翻译过程正在进行。尤其是酵母细胞中的跨内质网膜和细菌中跨细胞膜（认为和内质网在进

化上相关；图 12-4）的翻译后蛋白质易位十分普遍。为了在翻译后易位中仍发挥作用，易位体需要辅助蛋白将多肽链送入孔道引发易位（图 12-45）。在细菌中，一种易位动力蛋白，SecA ATP 酶，结合在易位体的胞质面，通过 ATP 水解促使动力蛋白发生周期性的构象转变。每次一个 ATP 水解时，SecA 蛋白的一段就插入易位体孔道中，将一小段转运蛋白推入孔道中。这种棘轮机构的结果是使 SecA 蛋白运送多肽链完全通过膜。

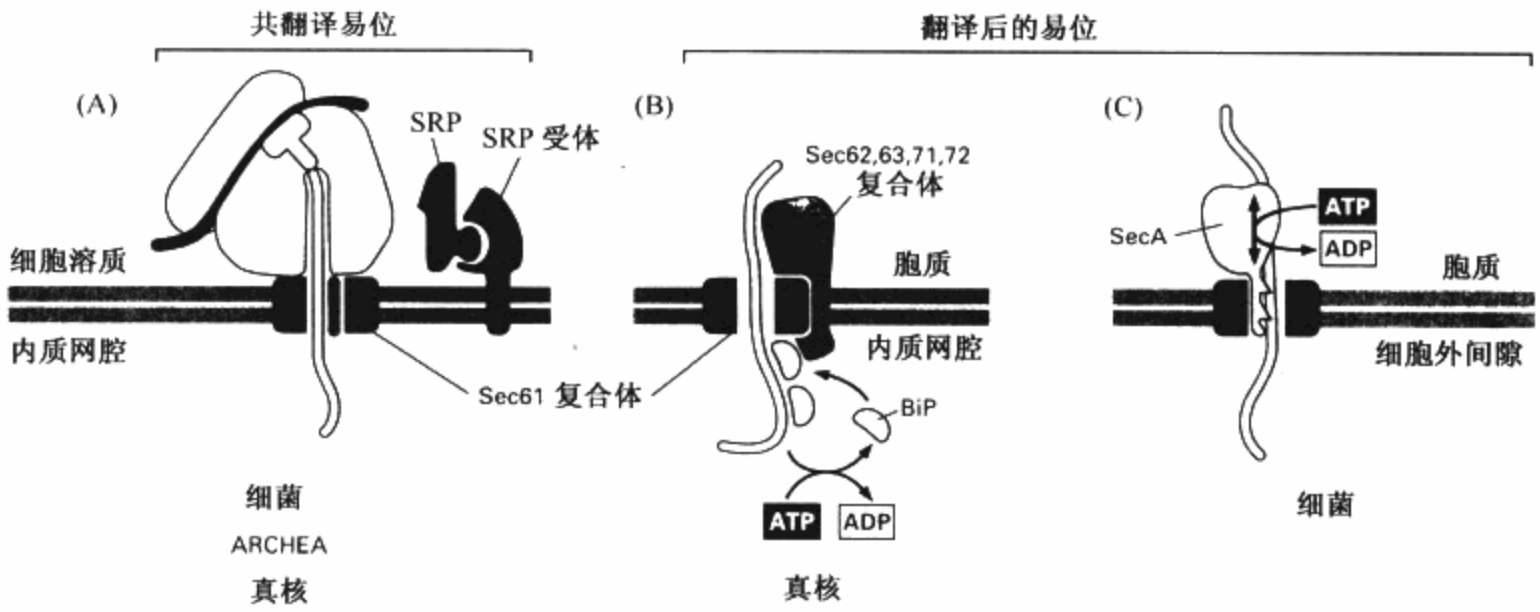


图 12-45　结构相似的易位体驱动蛋白质易位的 3 种方式

（A）协翻译易位。核糖体被 SRP 和 SRP 受体带至膜，和 Sec61 蛋白易位体形成紧密结合。多肽链与合成时一样呈线性经过膜。由于跨膜是正在延伸的多肽链的唯一途径，这一过程不需要额外的能量。（B）真核细胞中的翻译后易位。一个由 Sec62、Sec63、Sec71 和 Sec72 蛋白组成的附加复合体结合在 Sec61 易位体上。当易位多肽链进入内质网腔时，附加复合体将 BiP 分子沉积在其上。由 ATP 驱动的 BiP 分子结合释放循环将蛋白质拉入内质网腔。这一机制和图 12-28 中阐述的线粒体输入的热棘齿模型十分相似。（C）细菌中的翻译后易位。合成完成的多肽链通过 SecA ATP 酶的作用从质膜上易位体的胞质面进入易位体。ATP 水解驱动的构象转变使 SecA 发生注射器活塞样的运动，每一次反应循环都将蛋白质多肽链上的 20 个氨基酸推过易位体孔道。经过叶绿体类囊体膜的蛋白质易位的 Sec 路径和这一机制相似［图 12-30(B)］。SRP 和 SRP 受体在所有生物体中都存在，而 Sec61 易位体则并非如此。SecA 仅发现存在于细菌中，而 Sec62、Sec63、Sec71 和 Sec72 蛋白仅存在于真核细胞中

真核细胞使用一组不同的和 Sec61 复合体相连的辅助蛋白。这些跨内质网膜的蛋白质通过内质网膜腔面上的一个小结构域使一个 hsp70 样的伴侣蛋白（称作 BiP，即结合蛋白）结合在刚从孔道中进入内质网腔的多肽链上。BiP 的结合分离周期驱动了单向易位，这和前面提到的线粒体 hsp70 将蛋白质拉过线粒体膜的过程很相似。

通过翻译后机制运送进入内质网的蛋白质首先被释放进入细胞质基质中，和前面讨论的线粒体、叶绿体蛋白一样，它们通过和伴侣蛋白的结合防止自身折叠。在这种情况下，易位在没有核糖体密封孔道的条件下发生，多肽链究竟如何在没有离子和其他分子外漏的情况下通过易位体孔道仍然是一个未解之谜。

易位后大部分可溶性蛋白质的内质网信号序列被切除

我们已经知道在叶绿体和线粒体中，一旦前体蛋白通过膜其信号序列即被切除。相似地，内质网 N 端信号序列也被内质网膜腔面上的信号肽酶切除。但是信号序列本身不足以引发信号肽酶的切除，而是需要一个与之邻接的酶切位点以被信号肽酶特异性识别。我们在下文将会看到多肽链内部而非 N 端的内质网信号序列不含有这样的识别位点因而不会被切除；它们可以帮助跨膜蛋白在易位完成后停留在类脂双层膜中。

可溶性蛋白质的 N 端内质网信号序列具有双重信号功能。它引导蛋白质至内质网膜，又作为**起始转移信号**（起始穿膜肽）打开孔道。即使已被信号肽切除，信号序列仍被认为和易位体相连，以至于剩余部分蛋白质在线形穿过膜时形成一个大环袢。一旦蛋白质的 C 端穿过了膜，易位蛋白就被释放到内质网腔中（图 12-46）。然后信号序列从孔道中被释放出来被内质网中的其他蛋白酶迅速降解为氨基酸。

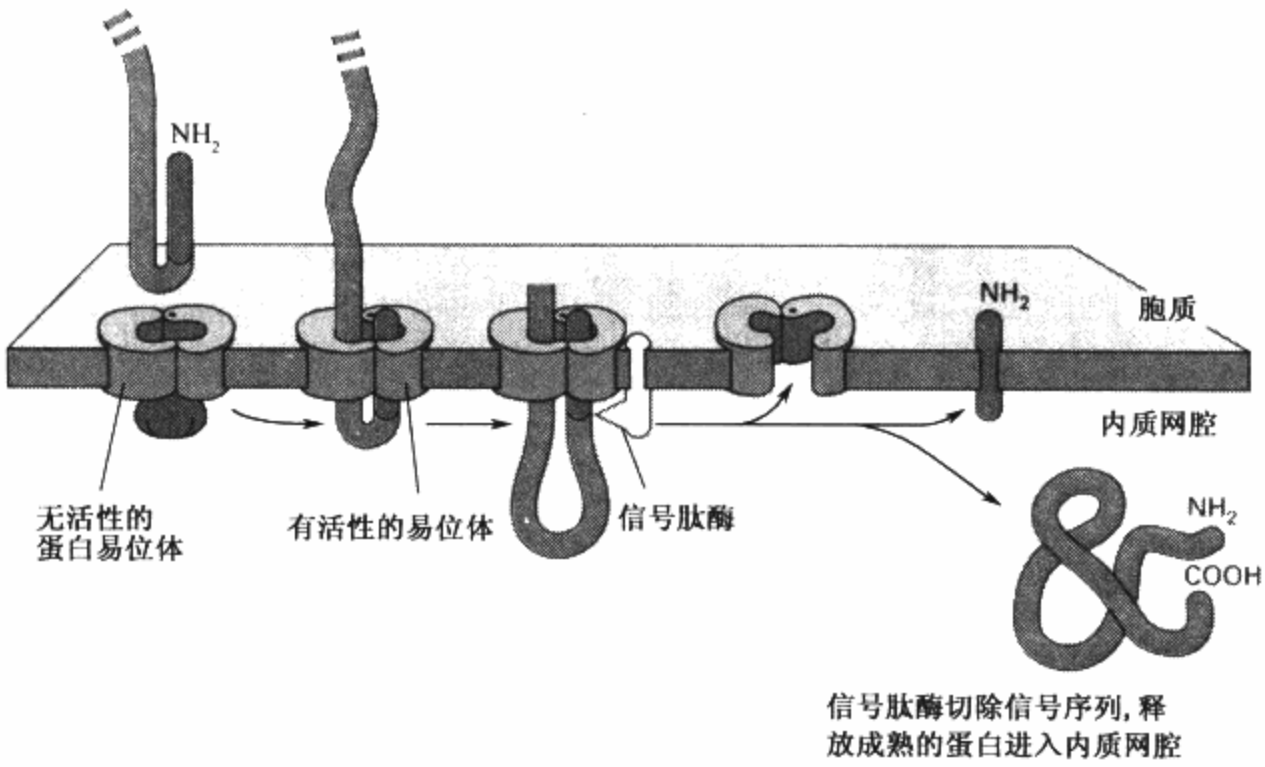


图 12-46　可溶性蛋白质经过内质网膜易位的模型

当和内质网信号序列结合时（作为起始转移信号），易位体孔道打开，使多肽链以环袢状的形式穿过类脂双层膜。当蛋白质完成易位后，孔道关闭，但是易位体向类脂双层膜内部侧向打开，使疏水性信号序列进入双层膜并在此处快速降解（在这幅图及后续三幅图中，核糖体被省略以使其更加清晰）

当和易位体孔道结合时，信号序列不仅和形成孔道壁的 Sec61 复合体接触，还和内质网膜的疏水性脂质部分接触。化学交联实验中信号序列和脂类疏水链的共价连接证明了这一点。为了将信号序列释放于膜中，易位体必须可以单侧打开。因此易位体在两个方向都有门控效应：当两侧都打开时就形成一个跨膜孔道可以使蛋白质的亲水性部分穿过类脂双层膜，当单侧打开时可以使蛋白质的疏水性部分插入类脂双层膜中。这种单侧门控机制对于跨膜蛋白插入类脂双层膜十分重要，我们将在下面进行讨论。

在单次跨膜蛋白中，唯一的内质网内在信号序列以跨膜 α 螺旋的形式保留在类脂双层膜中

最终停留在膜中的蛋白质易位过程比可溶性蛋白质要复杂许多，因为一部分多肽链通过了类脂双层膜而另一部分则未通过。尽管如此，所有膜蛋白的插入模式都可以看作可溶蛋白质进入内质网腔的一系列步骤的变形。下面我们先从**单次跨膜蛋白**（图 10-17）插入内质网的 3 种模式开始来讨论这一问题。

一种最简单的情况是 N 端信号序列引发易位，此步骤和可溶性蛋白质相同，但是在整条多肽链完成跨膜前，其中的疏水性片段会中止易位过程。在内质网信号序列（起始转移信号）被易位体释放并被切除后，这些**终止转移信号**把蛋白质锚定在膜上（图 12-47）。终止转移信号通过单侧门控机制被转运到类脂双层膜中，并以跨膜 α 螺旋片段的形势在此处停留下来，

其 N 端在膜的腔面，C 端在膜的胞质面。

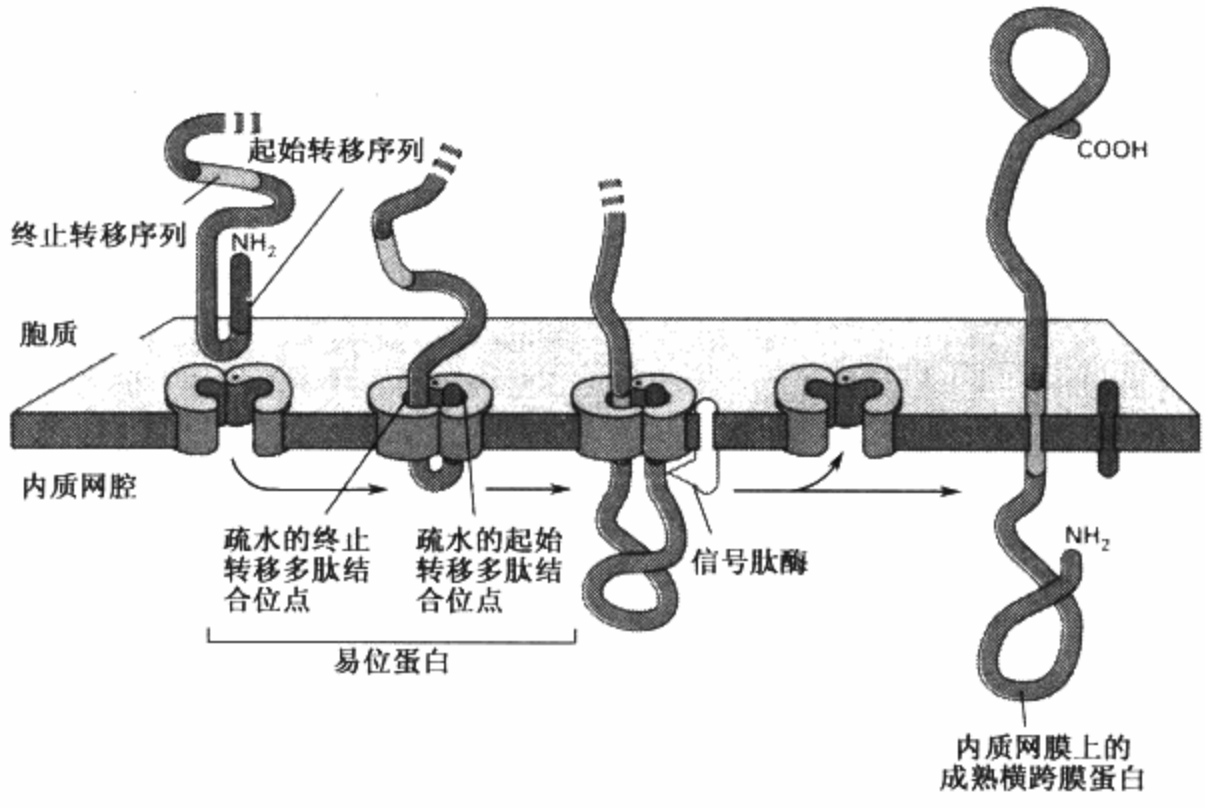


图 12-47　拥有可切除内质网信号序列的单次跨膜蛋白如何整合到内质网膜上

对于这个假想蛋白质，协翻译易位过程由 N 端作为起始转移信号的内质网信号序列（红色）引发，如图 12-46 所示。除了这个起始转移序列以外，蛋白质还含有一个终止转移序列（橙色）。当终止转移序列进入易位体并和一个结合位点相互作用时，易位体就改变构象并侧向卸载蛋白进入类脂双层膜

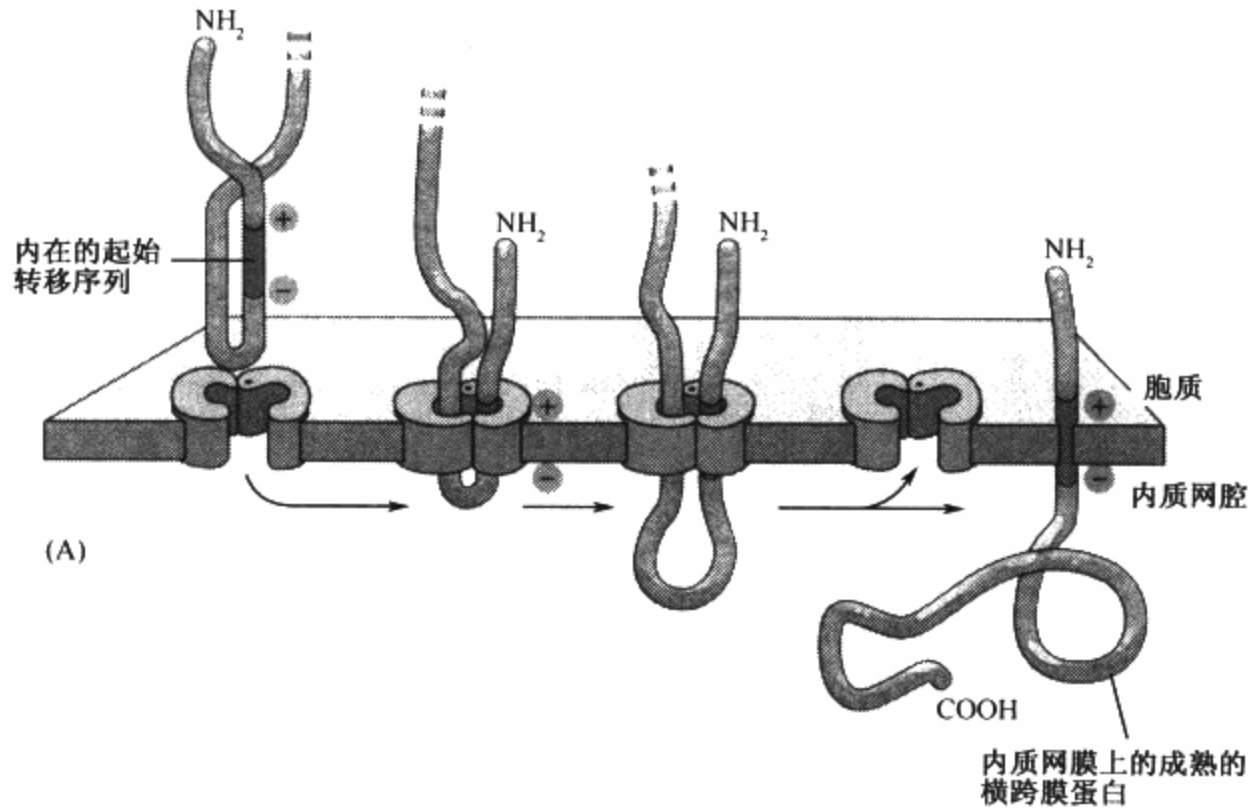
在其他两种情况中，信号序列是内在的而不是处于蛋白质的 N 端。但是和 N 端内质网信号序列一样，这些内部转移序列也被 SRP 识别，将正在合成蛋白质的核糖体带到内质网膜上并作为起始转移信号引发蛋白质易位。被易位体释放后，内在起始转移信号仍以单次跨膜 α 螺旋的形式保留在类脂双层膜中。

内在起始转移信号序列可以从 2 个方向和易位装置结合，插入的起始转移信号序列的方向决定了哪一段蛋白质片段（在起始转移信号之前的还是之后的）穿过膜进入内质网腔。一种情况是膜蛋白的 C 端在内质网膜腔面［图 12-48(A)］，另一种情况是 N 端在腔面［图 12-48(B)］。起始转移信号序列的方向如图中所示，由其近旁的带电氨基酸表明。

起始转移信号和终止转移信号的组合决定了多次跨膜蛋白的拓扑学结构

多次跨膜蛋白的多肽链来回多次穿过类脂双层膜（图 10-17）。内部转移序列作为这种蛋白质的起始转移信号引发易位，直到遇到终止转移信号时才停止易位。在两次跨膜蛋白中，多肽链随后即可以被释放于类脂双层膜中（图 12-49）。而更复杂的多次跨膜蛋白拥有很多疏水性的跨膜 α 螺旋，因此就需要第二个起始转移信号再次引发后续多肽链的易位过程，直到下一个终止转移信号引起多肽链的释放，或者引发后续的起始穿膜和终止穿膜序列行使功能（图 12-50）。

一个给定的疏水性信号序列是作为起始转移序列还是作为终止转移序列是由它们在多肽链中的位置决定的，因为如果通过重组 DNA 技术改变它们在蛋白质中的位置，功能也将被相应改变。因此，起始转移序列和终止转移序列的差异主要来自它们在被合成的多肽链中的相对位置。SRP 似乎是从伸展多肽链的 N 端开始，向 C 端依次搜寻疏水片段，即顺着蛋白质被合成的方向。通过识别从核糖体中出现的第一个适宜的疏水性片段，SRP 设置“可读框”：



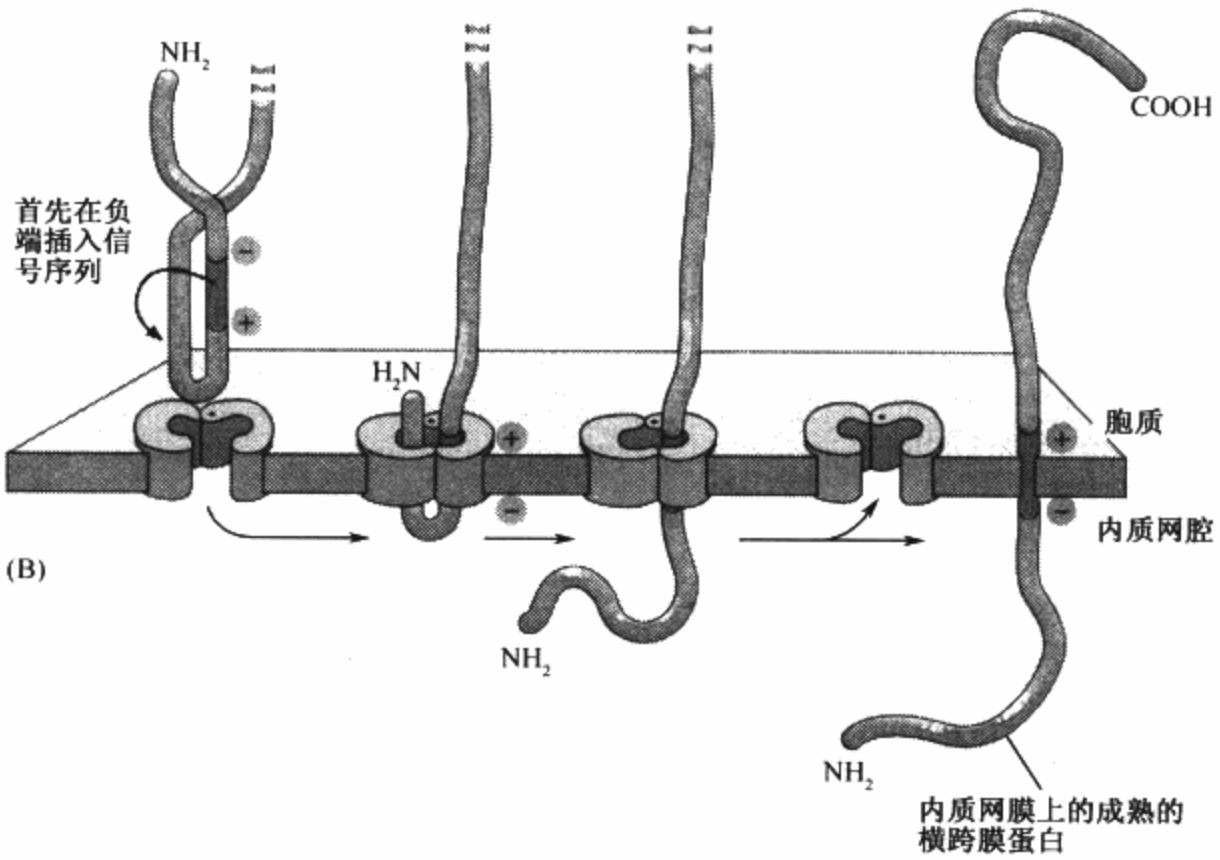


图 12-48　具有内在信号序列的单次跨膜蛋白插入内质网膜

对于这种假想蛋白质，内质网内在信号序列作为起始转移信号结合易位体，其带正电荷的一端仍留在细胞质基质中。（A）如果起始转移序列疏水性核心之前相邻的部分比之后相邻的部分具有更多的正电荷氨基酸，则起始转移序列插入易位体的方向正如此图中所示。相对于信号序列的 C 端一侧会穿过膜。（B）如果起始转移序列疏水性核心之后相邻的部分比之前相邻的部分具有更多的正电荷氨基酸，则起始转移序列插入易位体的方向正如此图中所示。相对于信号序列的 N 端一侧会穿过膜。因为易位过程在起始转移序列离开核糖体之前不能发生，因此（B）中蛋白质 N 端部分的易位在这一部分完全合成后就可发生了。注意有两种方式可将 N 端定位于内质网腔的单次跨膜蛋白插入膜：一种如图 12-47 所示，另一种如此图（B）所示

如果易位开始，下一个适宜的疏水性片段就会被识别作终止转移序列，引起两个序列间的多肽链线性穿过膜。相似的搜寻过程会一直持续直到所有的蛋白质疏水区域都被插入膜中。

由于膜蛋白总是按照这种程序从内质网的胞质面插入膜，所有相同的多肽链会在类脂双

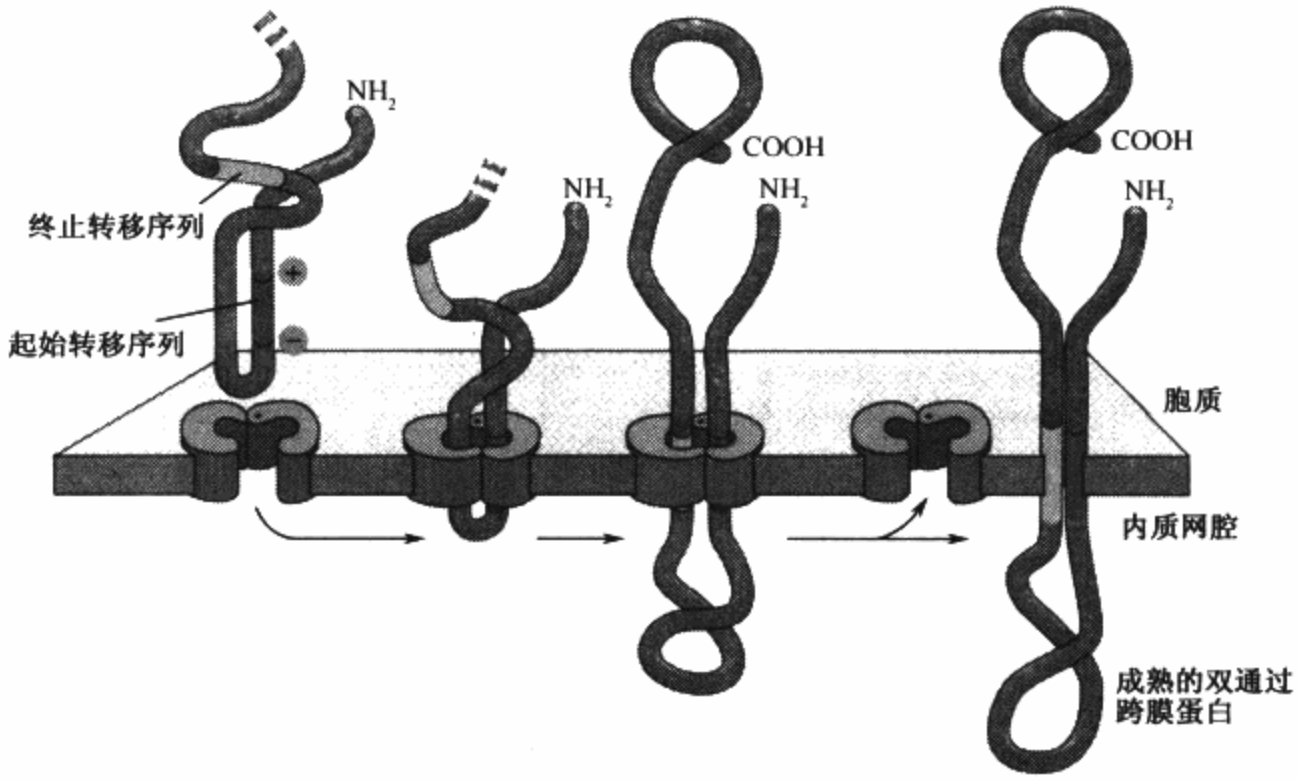


图 12-49　具有内在信号序列的 2 次跨膜蛋白插入内质网膜

对于此中假想蛋白质，内质网内在信号序列作为起始转移信号（图 12-48）并引发蛋白质 C 端部分的易位。当终止信号序列进入易位体后，易位体将蛋白质侧向卸载至膜中

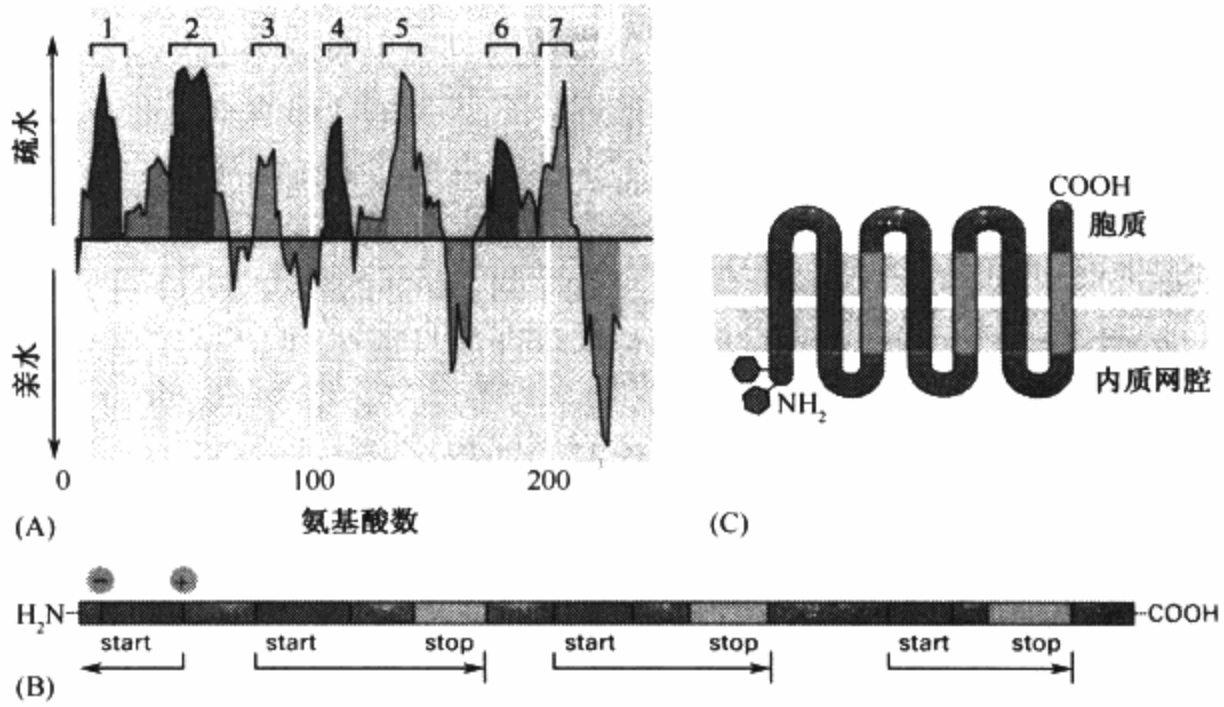


图 12-50　多次跨膜蛋白视紫红质插入内质网膜

视紫红质是哺乳动物视网膜感光视杆细胞中一种光敏蛋白（讨论见第 15 章）。（A）疏水性图显示视紫红质中有 7 个短的疏水性结构区。（B）最靠近 N 端的区域作为起始转移序列引起蛋白质前面的 N 端部分穿过内质网膜。后续的 6 个疏水性序列交替作为起始信号序列和终止信号序列行使功能。（C）最终插入内质网膜的视紫红质的 N 端在内质网腔中，C 端在细胞质基质中。蓝色的六边形达标共价连接的寡糖。箭头表示插入易位体的蛋白质部分

层膜中有相同的方向。由于蛋白质在内质网膜一侧所暴露的结构域和在另一侧所暴露的结构域不同，这就会引起内质网膜的不对称。膜在出芽和融合，将内质网蛋白质运送至其他细胞膜系统的过程中始终维持着这种不对称性（讨论见第 13 章）。这样，新合成的蛋白质插入内质网膜中的方向决定了这个蛋白质在其他所有的膜系统上的方向。

当蛋白质被从膜上离解下来然后被用于重组人工脂囊泡时，经常会出现蛋白质正确朝向

和相反朝向随机混杂的结果。因此，在细胞的膜系统上观察到的蛋白质不对称性似乎不是蛋白质的内在性质，而仅仅是由蛋白质从细胞质基质插入内质网膜这一过程造成的。

易位多肽链在糙面内质网腔中折叠组装

很多蛋白质只是暂时停留于内质网腔中，等待进入其他目的地；其他蛋白质则以高浓度留在内质网中。这些**内质网驻留蛋白**都在其 C 端含有一个 4 个氨基酸的**内质网驻留信号**，负责将这些蛋白质保留在内质网中（表 12-3，讨论见第 13 章）。一些驻留蛋白质作为催化剂帮助很多易位进入内质网的蛋白质正确折叠和组装。

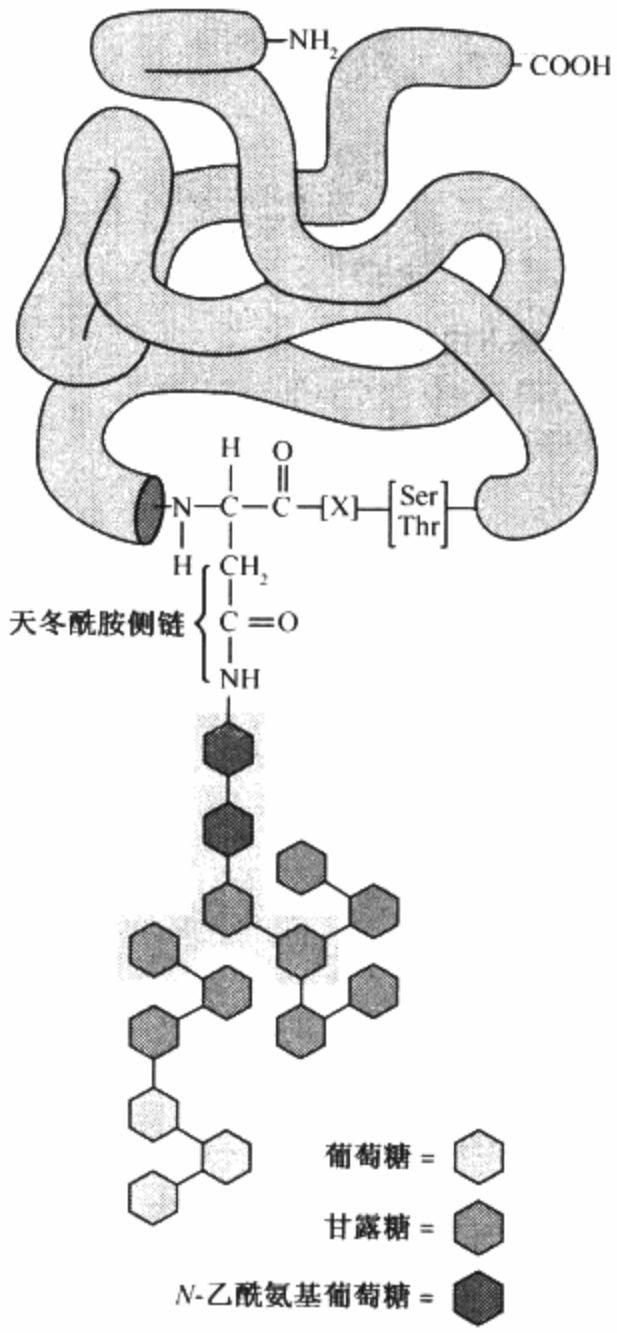


图 12-51　天冬酰胺-连接（*N*-连接）寡糖前体被加载到大多数粗面内质网膜蛋白上

灰框中的五糖形成这个寡糖的核心区域。对于很多糖蛋白来说，在高尔基体中发生的寡糖剪切过程中只有核心糖能被保留下来。在 Asn-X-Ser 和 Asn-X-Thr 序列中（X 代表除脯氨酸以外的任何氨基酸）只有天冬酰胺被糖基化。这两种序列在糖蛋白中出现的频率远远小于未糖基化的细胞质基质蛋白；显然在蛋白质进化的进程中存在不利于这些序列的选择压，这可能是由于过多位点的糖基化会影响蛋白质的正常折叠

一种重要的驻留蛋白质是蛋白质二硫键异构酶（PDI），它催化半胱氨酸上的自由巯基（—SH）氧化形成二硫键（—S—S—）。几乎所有在分泌和胞吞途径中暴露于细胞外间隙和细胞器腔内的蛋白质结构域上的半胱氨酸都形成了二硫键；但是暴露于细胞质基质中的半胱氨酸则不能形成二硫键，因为这里是一个还原环境。

另一种内质网驻留蛋白质是伴侣蛋白 **BiP**。我们已经讨论了 BiP 如何将转录后的蛋白质从内质网易位体拉入内质网。和其他伴侣分子一样，BiP 识别那些错误折叠的蛋白质和未最终组装成寡聚复合体的蛋白质。这样，它就会结合那些本应埋在正确折叠或完成装配的多肽链内部的氨基酸序列。一个 BiP 结合位点的例子是一股本应埋于 β 片层中的疏水性、亲水性氨基酸交替序列。BiP 的结合不但防止蛋白质聚集，还可以帮助蛋白质保留在内质网中（保证其不进入高尔基体或分泌途径）。和结合细胞质基质中的伸展蛋白质并促进它们输入线粒体及叶绿体的 hsp70 家组相似，BiP 水解 ATP 以帮助蛋白质折叠，并将转录后的蛋白质输入内质网。

大多数糙面内质网中合成的蛋白质通过加 *N*-连接寡糖而被糖基化

将糖链共价连接在蛋白质上是内质网的主要生物合成功能之一。大多数内质网形成的可溶和膜结合蛋白（包括最终运送至高尔基体、溶酶体、细胞膜或者细胞外间隙的蛋白质）都被糖基化。相比较而言，细胞质基质中很少有蛋白质被糖基化，发生了糖基化的蛋白质修饰也极为简单，一般是将 *N*-乙酰葡糖胺单链加在蛋白质的丝氨酸和苏氨酸残基上。

在理解**蛋白质糖基化**过程上的重大进展是发现已经形成的寡糖前体（由 *N*-乙酰葡糖胺、甘露糖和葡萄糖组成，总共含有 14 个单糖）整体被转移到内质

网中的蛋白质上。由于这个寡糖被转移到精氨酸侧链的 NH_2 官能团上，因此被称为 *N*-连接或精氨酸连接（图 12-51）。转移由一种称作寡糖基转移酶的膜结合酶催化，这种酶的活性位点在内质网膜的腔面；这解释了细胞质基质蛋白为什么没有这种形式的糖基化。寡糖前体被一种称为**多萜醇**的特殊脂类分子保留在内质网膜上，在精氨酸残基经过蛋白质易位过程出现在内质网腔中时，寡糖前体就立即通过一步酶催化反应转移到目标精氨酸上（图 12-52）。由于大部分蛋白质是协同翻译输入内质网的，所以 *N*-连接寡糖几乎总是在蛋白质合成过程中加到蛋白质上的。

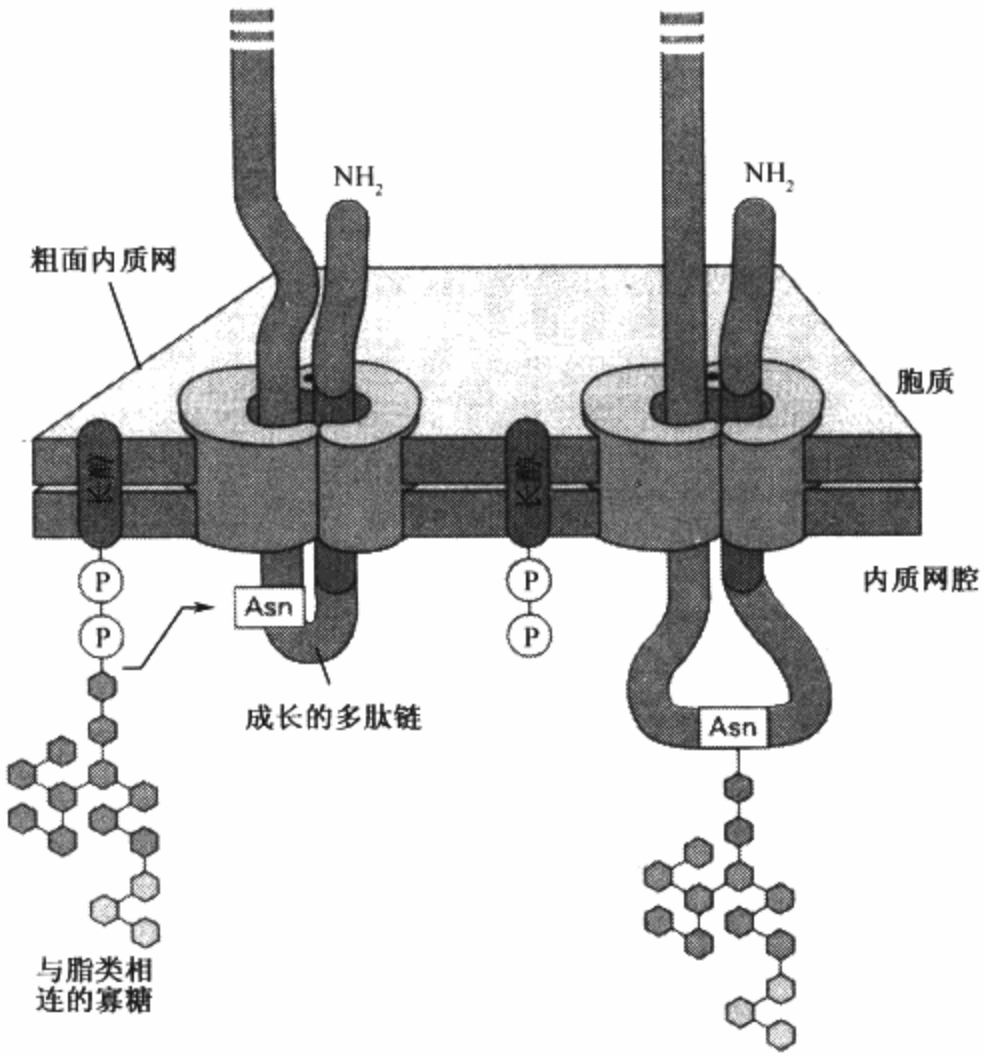


图 12-52　粗面内质网中的蛋白质糖基化

多肽链几乎一进入内质网腔其目标天冬酰胺就被糖基化。图 12-51 中显示的寡糖前体以一个完整的单位被连接到天冬酰胺上，这一反应由膜结合寡糖基转移酶催化。每一个内质网蛋白质易位体都连接有一份这种酶以及信号肽酶（未显示核糖体以便更清晰）

寡糖前体通过一个高能磷酸键连接在多萜醇脂类上，为图 12-52 所示的糖基化反应的启动提供了活化能。在转移到蛋白质上以前，整个寡糖前体是一个单糖一个单糖地加到膜结合的脂类分子上的。每一个单糖分子先在细胞质基质中被激活形成核苷酸糖中间物，然后再依次将其糖环部分和脂类结合（直接或间接地）。当这一过程进行到一半时，脂-连接寡糖链会从内质网膜的胞质面翻转至腔面（图 12-53）。

所有成熟糖蛋白 *N*-连接寡糖结构的多样性都来源于最初寡糖前体的修饰。在内质网中，大部分糖蛋白的寡糖链被迅速移除 3 个葡萄糖（图 12-51）和 1 个甘露糖。我们将在下文中讨论葡萄糖剪切的重要意义。这种寡糖链剪切和处理的过程在高尔基体中还将继续，我们将留到第 13 章进行讨论。

到目前为止，*N*-连接寡糖始终是发现的所有糖蛋白中最普遍的寡糖。有时，寡糖也会和

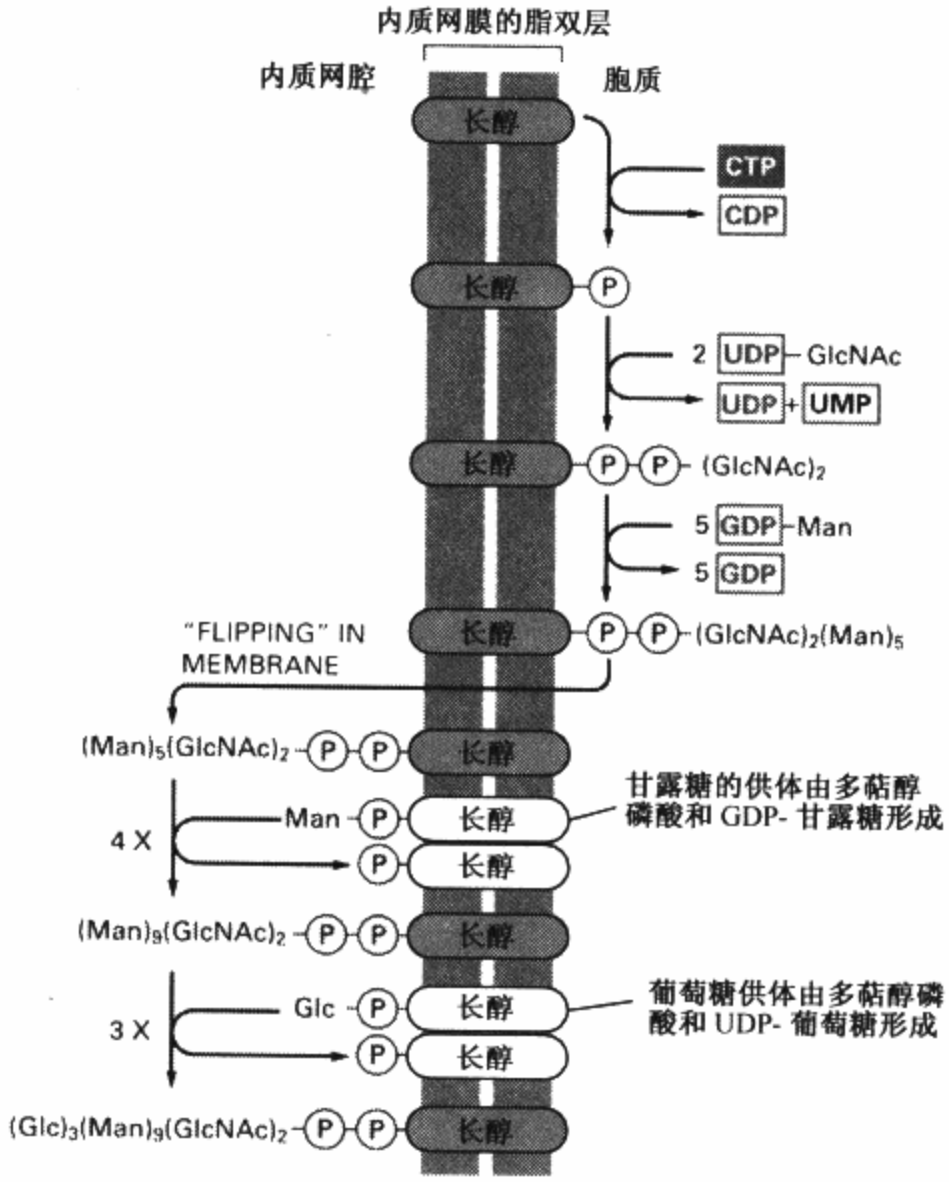


图 12-53　粗面内质网中脂-连接寡糖链的合成

寡糖链是由单糖依次装载到多萜醇脂（lipid dolichol）载体上形成的（一种聚类异戊二烯，注释 2-5）。多萜醇长且具疏水性：其 22 个五碳单位可以穿越类脂双层膜 3 次以上，因此连接其上的寡糖被牢牢地固定在膜上。最靠近多萜醇的第一个糖环以一个焦磷酸盐桥和多萜醇连接。这个高能键最终活化寡糖从脂类转移到刚刚伸入粗面内质网腔面的多肽链天冬酰胺侧链上。如图所示，寡糖的合成先从内质网的胞质面起始，当 $(Man)_5(GlcNAc)_2$ 脂类中间产物在运载体蛋白的帮助下进行了跨膜翻转后，合成过程将继续在腔面进行。所有后续在内质网腔面发生的糖基转移反应都包括多萜醇-P-葡萄糖和多萜醇-P-甘露糖的糖环传递；这些活化状态的脂连接单糖是由多萜醇磷酸和 UDP-葡萄糖或 GDP-甘露糖在内质网胞质面合成，再经过跨膜翻转进入内质网腔面的。GlcNAc＝*N*-乙酰葡糖胺；Man＝甘露糖；Glc＝葡萄糖

丝氨酸、苏氨酸或者羟赖氨酸侧链上的羟基相连。这种 *O*-连接寡糖在是高尔基体中通过一种至今尚未明确的途径形成的。

寡糖是标记蛋白质折叠状态的标签

关于为什么进入内质网的蛋白质会发生如此普遍的糖基化曾引发了长期的争论。一个特别令人费解的现象是一些蛋白质需要 *N*-连接糖基化来帮助自身在内质网中正确折叠，但这似乎和寡糖链连接于蛋白质表面的精确位置无关。蛋白质折叠过程中糖基化作用的线索来自于两种内质网伴侣蛋白的研究。由于两种蛋白质的活性都离不开 Ca^{2+}，因此它们分别被称作**钙连接蛋白**和**钙网织蛋白**。这些伴侣蛋白是凝集素，结合在未完成折叠的蛋白质的寡糖链上以保证它们驻留在内质网中。和其他伴侣蛋白相似，它们防止未完成折叠的蛋白质发生不可逆的聚集。钙连接蛋白和钙网织蛋白也都促进未完成折叠的蛋白质和其他内质网伴侣分子结合，

后者会结合还未形成二硫键的半胱氨酸。

钙连接蛋白和钙网织蛋白识别含有单个葡萄糖末端的 *N*-连接寡糖，因此仅能结合那些原先 3 个葡萄糖中的 2 个被内质网糖苷酶切除的蛋白质。当第 3 个葡萄糖被移除时，蛋白质将从它的伴侣蛋白上解离下来离开内质网。

那么，钙连接蛋白和钙网织蛋白如何区分完成折叠的蛋白质和未完成折叠的蛋白质呢？答案来自另一种内质网酶——葡萄糖基转移酶。它可以持续向丢失了最后一个葡萄糖的寡糖上添加一个葡萄糖单元。但是，它仅向连接在未折叠蛋白质上的寡糖添加葡萄糖。这样，未折叠蛋白质发生连续多次葡萄糖剪切（通过糖苷酶）和添加（通过糖基转移酶）后仍保持对钙连接蛋白和钙网织蛋白的亲和性，直到它们达到完全折叠的状态（图 12-54）。

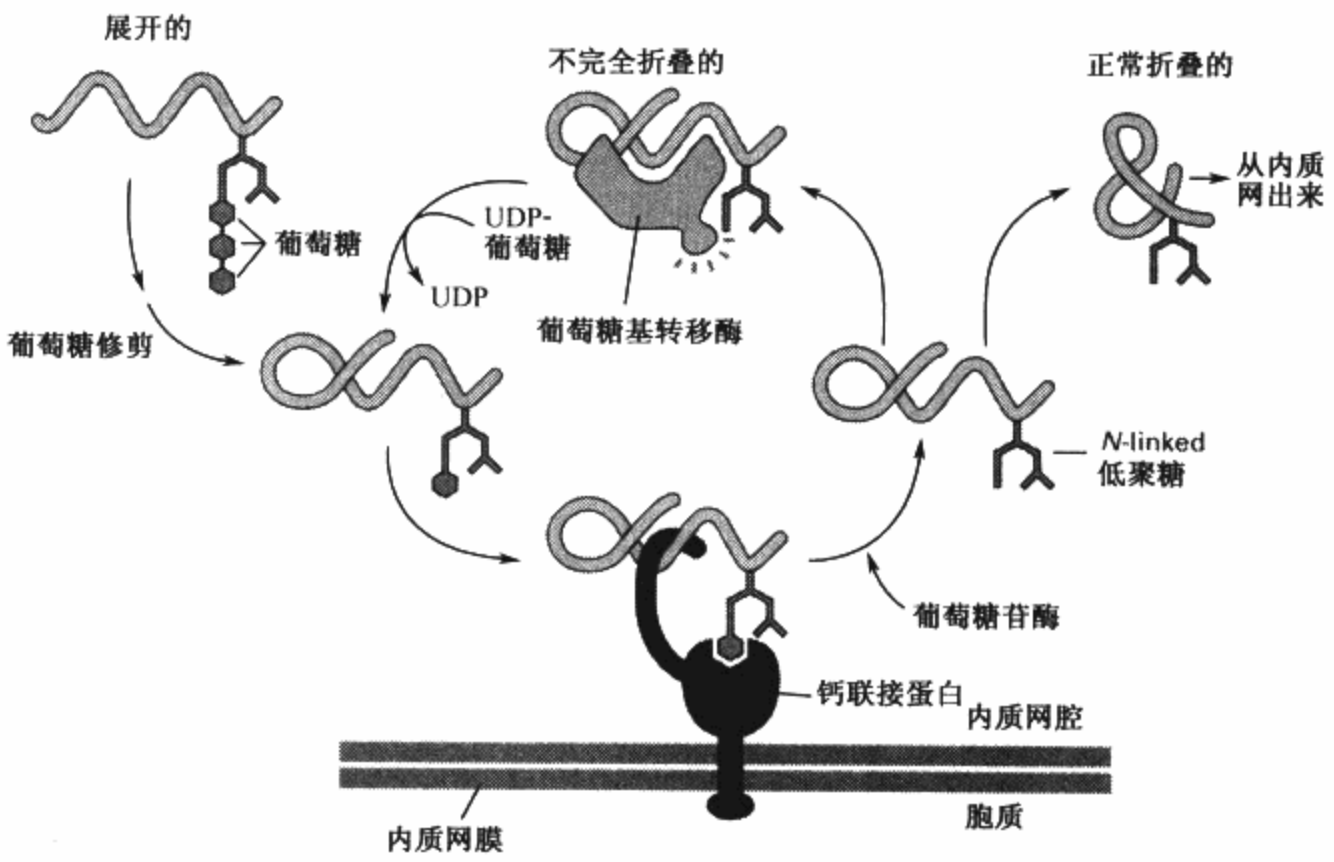


图 12-54　*N*-连接糖基化在内质网蛋白质折叠中的作用

内质网膜结合伴侣蛋白钙连接蛋白结合未完全折叠蛋白，但后者的 *N*-连接寡糖上必须含有一个葡萄糖尾。钙连接蛋白将这些蛋白质阻拦于内质网中。糖苷酶将葡萄糖尾切除，蛋白质即从钙连接蛋白中释放出来。葡萄糖基转移酶是决定蛋白质折叠正确与否的关键性酶：如果蛋白质尚未折叠完成，此酶会将一个新的葡萄糖从 UDP-葡萄糖上转移到蛋白质的 *N*-连接寡糖上，改变蛋白质对钙连接蛋白的亲和力，使其驻留在内质网中。这一过程循环重复直到蛋白质折叠完成。钙网织蛋白的功能与此相似，只是后者尾可溶性内质网居留蛋白。另一种内质网伴侣蛋白——Erp57（未显示）和钙连接蛋白以及钙网织蛋白协同作用将蛋白质驻留于内质网的过程中

未正确折叠的蛋白质被输出内质网在细胞质基质中降解

即使有伴侣蛋白的帮助，很多易位进入内质网的蛋白质分子都未能实现正确的折叠或寡聚体状态（有些蛋白质甚至超过 80%）。这样的蛋白质会被输出内质网进入细胞质基质被降解。这种反易位也称为转位，是通过蛋白质最初进入内质网时经过的相同易位体（Sec61 复合体）发生的。当然这一步需要其他蛋白质帮助易位体功能翻转。但是这些已经没有内质网信号序列的错误折叠蛋白是如何被识别或定位的目前尚不清楚。

一旦错误折叠蛋白进入细胞质基质，它们的寡糖即被切除。去糖基化被 *N*-聚糖酶（glycanase）催化，这种酶切断羰基和寡聚糖连接的精氨酸氨基之间的酰胺键移除寡聚糖。去糖

基化的多肽链迅速被内质网结合的泛素缀合酶泛素化，随即被送入蛋白酶体（讨论见第 6 章）中降解(图 12-55)。

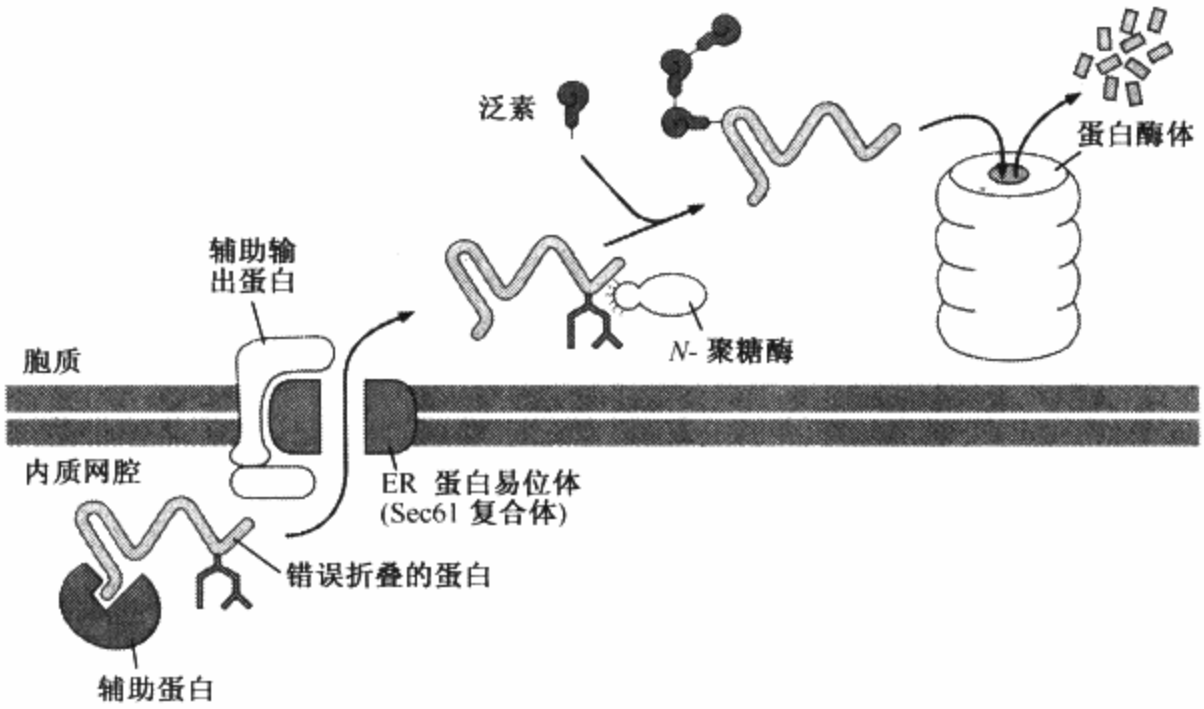


图 12-55　错误折叠的内质网蛋白质的输出和降解

内质网腔中错误折叠的可溶性蛋白质被易位回细胞质基质中，在此处它们被去糖基化、泛素化并被蛋白酶体降解。错误折叠的膜蛋白也以相同的方式降解。介导错误折叠蛋白输出的易位体和介导其输入的易位体类型相同，与易位体结合的辅助蛋白质使得易位体可执行输出易位

内质网中的错误折叠蛋白质可以引发蛋白质舒展反应

细胞严格控制着各个空间区域中所含的错误折叠蛋白的数量。例如，错误折叠蛋白在细胞质基质中的累积将引发热激反应（讨论见第 6 章），从而刺激编码细胞质基质伴侣分子的基因转录，以帮助重新折叠蛋白质。相似地，错误折叠蛋白在内质网中的累积将导致**蛋白质舒展反应**，包括增加编码内质网伴侣蛋白和内质网蛋白质降解酶类基因的转录。

细胞质和内质网中存在错误折叠的蛋白质的信号是如何进入细胞核的？酵母细胞中从内质网到细胞核的通路已经被研究得较为清楚。错误折叠蛋白先激活内质网的跨膜蛋白激酶，从而引发激酶的寡聚化和自磷酸化作用（细胞外生长因子以相似的方式激活它们的细胞膜受体，讨论见第 15 章）。内质网激酶的寡聚化引起该激酶分子上的内切核糖核酸酶结构域激活。这种核酸酶在 2 个位点切割细胞质基质 RNA 分子，除去 1 个内含子。分开的外显子被 RNA 连接酶连接，产生一条被剪接的 mRNA 在核糖体上翻译出基因调节蛋白。调节蛋白迁移进入细胞核并激活编码介导蛋白质舒展反应的蛋白质基因转录（图 12-56）。

一些膜蛋白获得共价连接的糖基磷脂酰肌醇锚点

正如第 10 章所述，一些细胞质基质蛋白质催化单一脂肪酸链或异戊二烯基和特定蛋白质的共价键结合。已经连接的脂类帮助引导这些蛋白质至细胞的膜系统。内质网膜酶可以催化一个相关的过程，将**糖基磷脂酰肌醇（GPI）锚点**共价连接在一些最终存在于细胞膜中的蛋白质的 C 端。这种连接在内质网腔中形成，同时蛋白质的跨膜片段被切除（图 12-57）。大多数细胞膜蛋白通过这种方法得到修饰。由于这些细胞膜蛋白质仅仅通过它们的 GPI 锚点附着

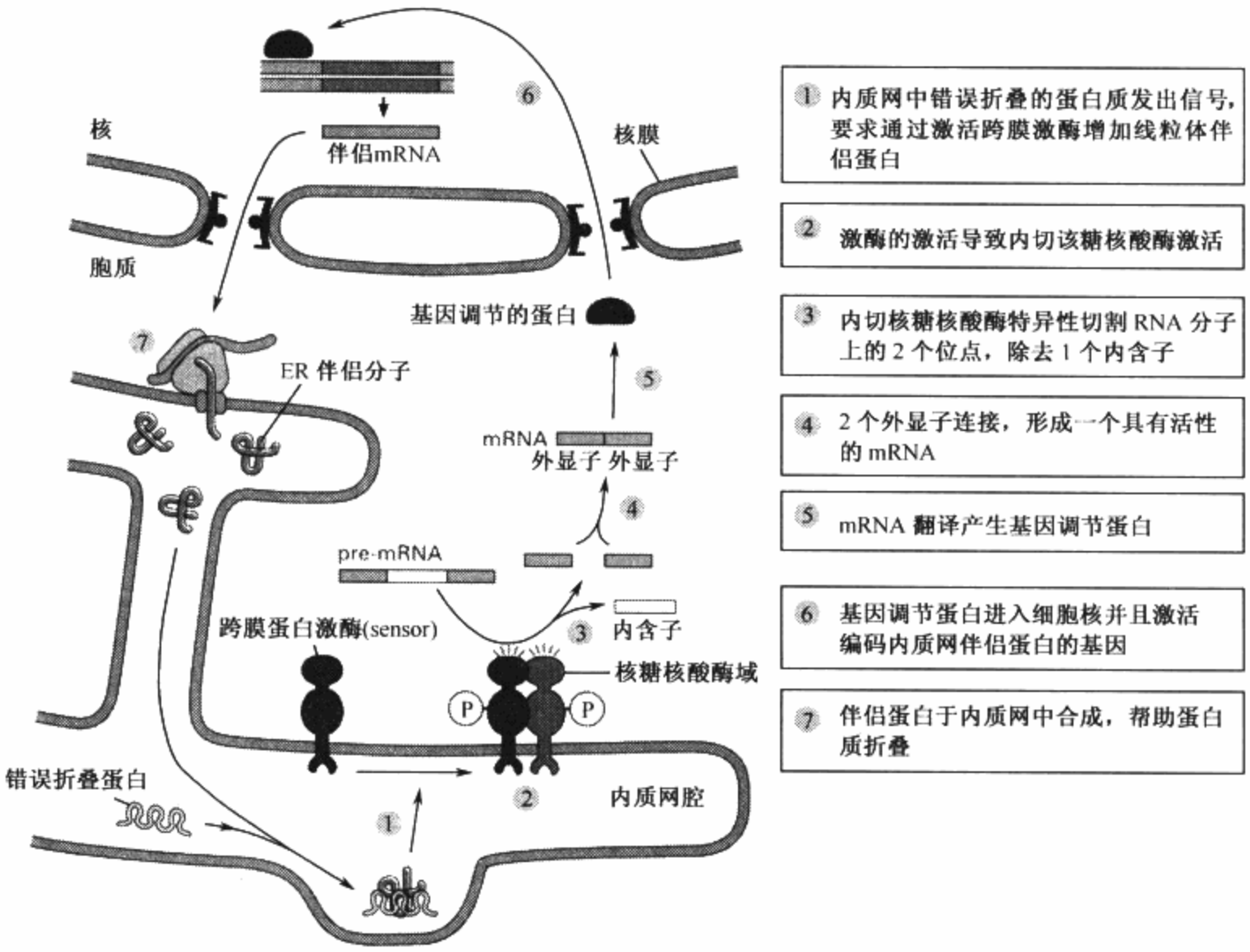


图 12-56　酵母中的未折叠蛋白应答

通过这种新奇的细胞内信号通路，内质网腔内累积的错误折叠蛋白向细胞核发出信号，激活一些基因，这些基因负责编码的蛋白质可以帮助细胞清除大量内质网中错误折叠蛋白

在细胞膜外表面，在受到激活细胞膜特异性磷脂酶的信号刺激后，原则上细胞膜蛋白可以以可溶性的形式被释放出去。例如，锥虫一旦受到免疫系统的攻击，就利用这一机制脱去 GPI-锚定表面蛋白。正如我们将在第 13 章中讨论的，GPI 也用来引导细胞膜蛋白进入脂筏以将这些蛋白质和其他膜蛋白分离开。

大多数类脂双层膜在内质网中组装

内质网膜合成几乎所有细胞膜需要的主要脂类，包括磷脂和胆固醇。其中一种主要的磷脂是卵磷脂（又叫蛋黄素），它由 1 个胆碱、2 个脂肪酸和 1 个甘油磷酸酯经 3 步反应形成（图 12-58）。每一步反应都被内质网上的酶催化，酶的活性位点朝向细胞质基质，因为所有需要的反应代谢物都在其中。因此，磷脂的合成仅发生在内质网膜的胞质面。在第一步中，酰基转移酶相继将 2 个脂肪酸加在甘油磷酸酯上产生磷脂酸，磷脂酸的水不溶性使它在合成后保留在类脂双层膜中。正是此步反应扩大了类脂双层膜的表面积。后两步反应决定了新合成的脂类分子的头部，进而决定了双层膜的化学结构，但不导致膜的净增长。另外两种主要的膜磷脂（磷脂酰乙醇胺和丝氨酸磷脂）以及一种次要的磷脂磷脂酰肌醇（PI）也是通过这种途径合成的。

由于磷脂的合成仅发生在内质网双层膜的胞质面一侧，因此需要一种机制将新合成的一

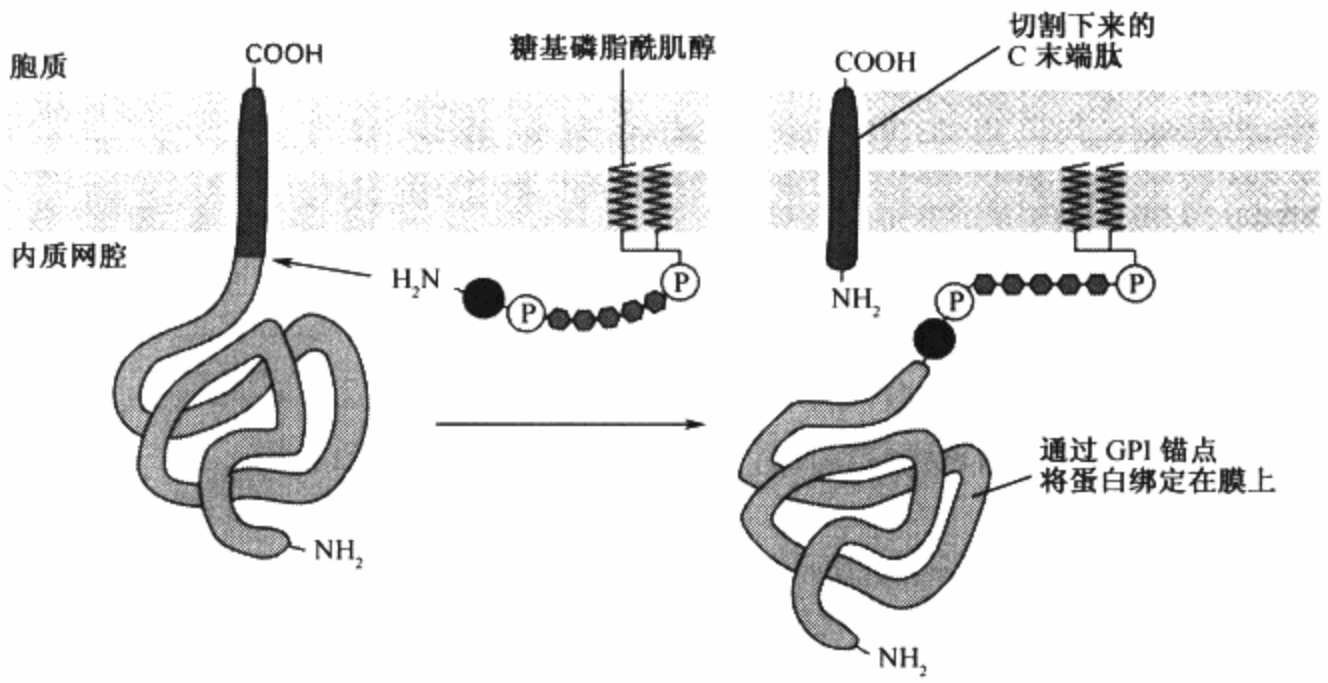


图 12-57　GPI 锚点和内质网内蛋白质的连接

在蛋白质合成刚刚完成时，蛋白质前体仍然通过一个疏水性的 15～20 个氨基酸的 C 端序列锚定在内质网膜上；而其余的蛋白质部分则在内质网腔中。不到 1min，内质网中的酶就会把蛋白质从膜结合的 C 端切割下来并将其新的 C 端和预装配好的 GPI 中间产物的氨基连接。这个修饰过程的特异性信号存在于疏水性的 C 端信号序列和内质网腔面与之毗邻的几个氨基酸中；如果这个信号加到其他蛋白质上，它们经过此修饰过程。由于具有共价连接的脂类锚点，蛋白质仍然结合在膜上。但是其氨基酸首次全部暴露于内质网腔面部分，最后暴露于细胞外面

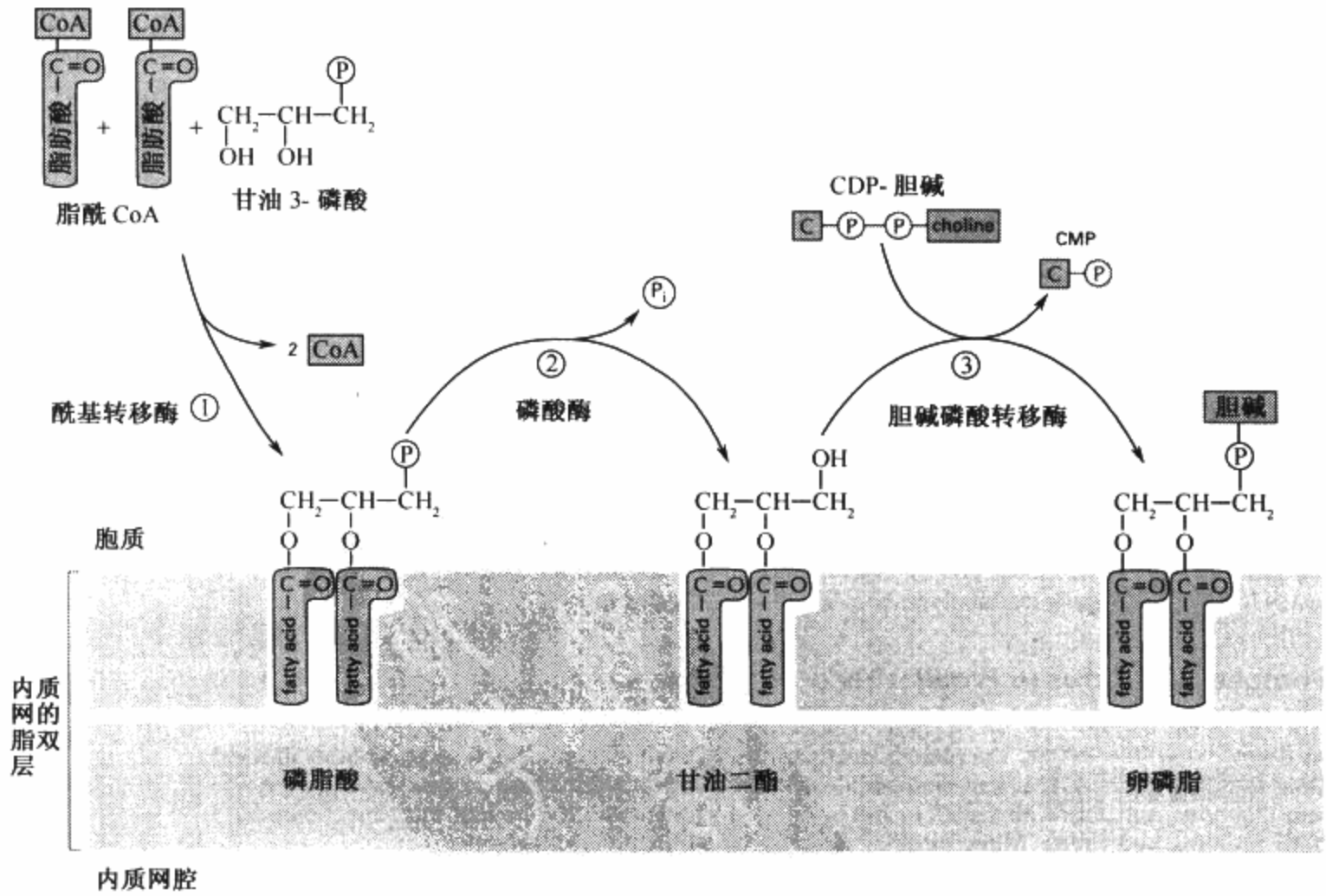


图 12-58　磷酸卵磷脂的合成

这种磷脂是由脂肪酸乙酰-辅酶 A、甘油 3-磷酸和胞嘧啶-二磷酸胆碱合成的

些磷脂分子转移到腔面来。在合成的类脂双层膜中，脂类分子并不是通过翻转来实现转移的。但是在内质网中，磷脂在几分钟之内就能达到膜两侧的平衡，这要比自发的翻转快将近100 000倍。这种快速的转膜运动被认为由一种称为 scramblase 的磷脂易位体介导，它能使类脂双层膜两叶的磷脂快速平衡（图 12-59）。这样，不同类型的磷脂就能在内质网的膜两叶平均分散。除了 scramblase，细胞膜还含有另一种类型的磷脂易位体，它属于 ABC 运载体家族（讨论见第 11 章）。这种翻转酶特异性地从类脂双层膜的细胞外叶移走含有自由氨基的磷脂（丝氨酸磷脂和磷脂酰乙醇胺），再利用 ATP 水解的能量使其定向翻转进入胞质叶。因此在翻转酶的维持下，细胞膜具有高度不对称的磷脂组成。

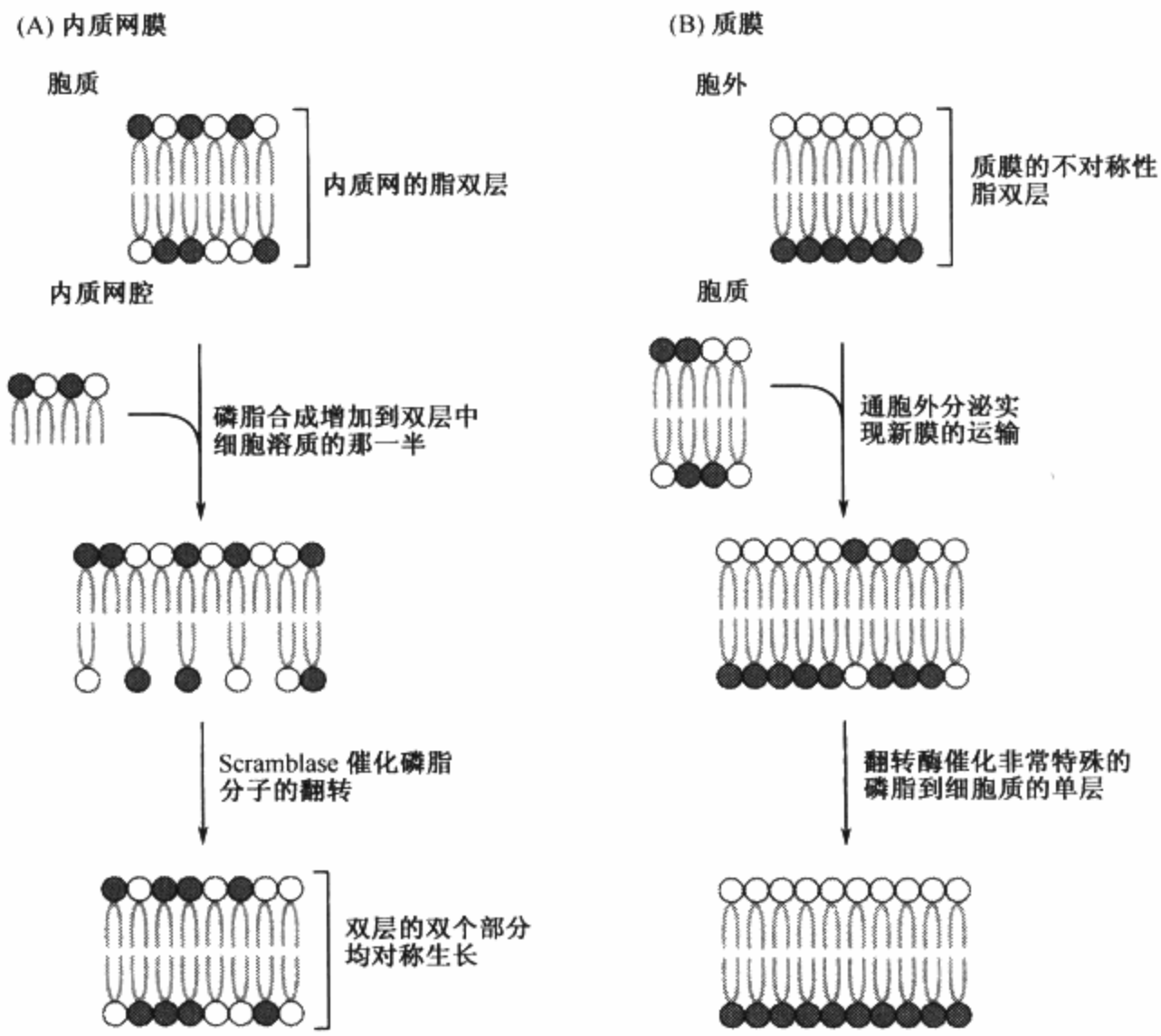


图 12-59　磷脂易位体在磷脂双层膜合成中的作用

（A）由于新的脂类分子只添加在双层膜胞质面层并且脂类分子并不会自发的从一层翻转至另一层，因而必须依赖膜结合磷脂易位体将脂类分子从胞质层转至腔面层，这样膜才能生长为双层形式。scramblase 不仅特异性地作用于特定的磷脂首基，并且在两层膜间平衡不同的磷脂。（B）由 ATP 水解驱动，质膜中首基特异性的翻转酶主动将丝氨酸磷脂和磷脂酰乙醇胺从胞外层定向翻转至胞质层，产生动物细胞质膜类脂双层膜的特征形不对称（图 10-14）。质膜中也具有 scramblase，以确保两层膜中拥有相同量的脂质；因此翻转酶的连续作用是维持磷脂对称所必需的

内质网也产生胆固醇和神经酰胺。丝氨酸和一分子脂肪酸发生聚合反应形成神经鞘氨醇；然后第 2 个脂肪酸分子加在神经鞘胺醇上形成神经酰胺。神经酰胺从内质网被运送至高尔基体，成为两种脂类合成的前体：将神经酰胺加上寡糖链形成糖鞘脂（糖脂），将磷酸卵磷脂的磷酸胆碱基团转移到神经酰胺上形成神经鞘磷脂。因此在膜合成的过程中，糖脂和神经鞘磷脂都形成得相对较晚。因为这两种脂类都是由高尔基体腔内的酶催化产生并且不是脂类易位体的底物，因此它们仅存在于类脂双层膜的非胞质叶。

磷脂交换蛋白帮助从内质网中将磷脂运送至线粒体和过氧化物酶体

正如第 13 章所述，细胞膜、高尔基体膜、溶酶体膜和胞内体膜都通过传递蛋白质和脂类的运输囊泡使自身的部分膜和内质网膜构成物质交流系统。但是线粒体、叶绿体以及过氧化物酶体则不属于这一系统，因此它们需要其他机制来输入生长必需的蛋白质和脂类。我们已经知道这些细胞器中的大部分蛋白质是从细胞质基质中转运的。此外，线粒体可以修饰它们输入的部分脂类，但不能从头合成它们；因此，它们的脂类必须从内质网中直接输入或通过其他细胞膜系统间接输入。不论上述哪一种情况，后一类细胞器都需要特殊的转运机制。

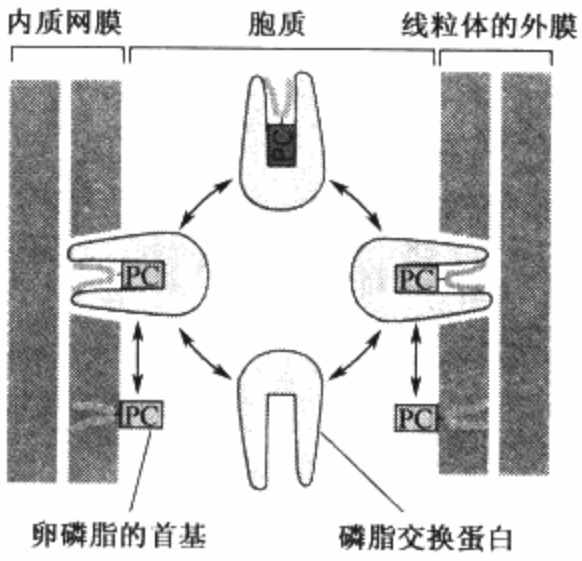


图 12-60　磷脂交换蛋白

由于磷脂不溶于水，它们在膜间的运输需要载体蛋白。磷脂交换蛋白是一种每次运载一个磷脂分子的水溶性蛋白质；它们可以从一膜上装载一个脂质分子再在另一膜上释放，以此在膜包被细胞器之间重新分配磷脂。从内质网到线粒体的磷酸卵磷脂（PC）的净转移可以不需要额外能量供给，因为内质网膜上（其生成处）的 PC 浓度高而线粒体外膜上的 PC 浓度低。预计线粒体外膜上具有脂质易位体以平衡外膜脂质双层膜的脂质，并且还具有一种机制促使脂质在线粒体外膜和内膜之间转移。但是这些假想的通路还未被发现

被称为**磷脂交换蛋白**的水溶性载体蛋白在膜之间传递磷脂分子。每一种交换蛋白仅特异性识别几种磷脂。它可以将磷脂分子从膜中“拉出”，再将埋藏于它的脂类结合位点的磷脂分子传递出去。当交换蛋白遇到其他膜时，会将结合的磷脂分子释放到新的类脂双层膜中（图 12-60）。丝氨酸磷脂被认为是通过这种方式输入线粒体的，并且在线粒体中脱羧产生磷脂酰乙醇胺（脑磷脂）。相比之下，卵磷脂被直接输入线粒体而无需任何修饰。

交换蛋白在所有膜系统间随机散布磷脂分子。原则上说，这种随机交换过程可以导致脂类从富含脂的膜到缺乏脂的膜的净运输。例如，这一过程可以使磷酸卵磷脂和丝氨酸磷脂分子从合成处内质网转运至线粒体或过氧化物酶体膜。在细胞质基质中，线粒体和过氧化物酶体可能是最缺乏脂类的细胞器，而交换过程可以满足它们对脂类的需要。在电子显微镜照片中，线粒体经常和内质网膜很近地并列排布在一起，线粒体特殊的脂类转运机制可能就发生在这个相互邻近的区域里。

小结

内质网网络结构是几乎所有细胞脂类的生产场所。除此之外，很大一部分细胞蛋白质的合成也发生在内质网的胞质面。所有最终将被分泌的蛋白质和所有最终保留在内质网、高尔基体、溶酶体、胞内体和细胞膜中的蛋白质都先从细胞质基质运送到内质网中。在内质网腔中，蛋白质折叠并寡聚体化，形成二硫键，添加 *N*-连接寡糖。*N*-连接糖基化用于指示蛋白质的折叠程度，所以蛋白质仅能在其正确折叠后才能离开内质网。不能正确折叠或寡聚化的蛋白质会被易位回到细胞质基质，并在此处去糖基化、泛素化，然后被蛋白酶体降解。如果错误折叠的蛋白质在内质网中过度累积，将引起蛋白质舒展反应，即激活细胞核中的适宜基因以帮助内质网应对累积的错误折叠蛋白。

只有那些携带特异性的内质网信号序列的蛋白质可以输入内质网。信号受体蛋白（SRP）识别信号序列，同时结合正在延长的多肽链和核糖体，将它们引导至粗面内质网的胞质面受体蛋白上。与内质网膜的结合引发易位过程，多肽链通过跨膜蛋白易位体的疏水孔道线性穿

过内质网膜形成一个环袢。

最终分泌、保留在内质网腔或者转移到其他细胞器的腔内的可溶性蛋白质会完全进入内质网腔。最终停留在内质网膜或其他细胞膜系统上的跨膜蛋白部分通过内质网膜，通过其多肽链上一个或多个跨膜α螺旋区将自身锚定在膜上。蛋白质的这些疏水片段在易位过程中作为起始转移信号和终止转移信号。当一条多肽链含有多个交替的起始转移信号和终止转移信号时，它会反复穿过类脂双层膜多次，形成一个多次跨膜蛋白。

内质网蛋白质插入和糖基化的不对称性为那些由内质网提供膜蛋白的细胞器的膜建立起方向。

参考文献

一般文献

Palade G (1975) Intracellular aspects of the process of protein synthesis. *Science* 189, 347–358.

The Compartmentalization of Cells

Blobel G (1980) Intracellular protein topogenesis. *Proc. Natl. Acad. Sci. USA* 77, 1496–1500.

Martoglio B & Dobberstein B (1998) Signal sequences: more than just greasy peptides. *Trends Cell Biol.* 8, 410–415.

von Heijne G (1990) Protein targeting signals. *Curr. Opin. Cell Biol.* 2, 604–608.

Warren G & Wickner W (1996) Organelle inheritance. *Cell* 84, 395–400.

The Transport of Molecules Between the Nucleus and the Cytosol

Adam SA (1999) Transport pathways of macromolecules between the nucleus and the cytoplasm. *Curr. Opin. Cell Biol.* 11, 402–406.

Arts G-J, Fornerod M & Mattaj IW (1998) Identification of a nuclear export receptor for tRNA. *Curr. Biol.* 8, 305–314.

Blobel G & Wozniak RW (2000) Proteomics for the pore. *Nature* 403, 835–836.

Gant TM & Wilson KL (1997) Nuclear assembly. *Annu. Rev. Cell Dev. Biol.* 13, 669–695.

Görlich D & Mattaj IW (1996) Nucleocytoplasmic transport. *Science* 271, 1513–1518.

Kaffman A & O'Shea EK (1999) Regulation of nuclear localization: a key to a door. *Annu. Rev. Cell Dev. Biol.* 15, 291–339.

Kalderon D, Roberts BL, Richardson WD & Smith AE (1984) A short amino acid sequence able to specify nuclear location. *Cell* 39, 499–509.

Lyman SK & Gerace L (2001) Nuclear pore complexes: dynamics in unexpected places. *J. Cell Biol.* 154, 17–20.

Mattaj IW & Conti E (1999) Snail mail to the nucleus. *Nature* 399, 208–210.

Politz JC & Pederson T (2000) Review: movement of mRNA from transcription site to nuclear pores. *J. Struct. Biol.* 129, 252–257.

Stoffler D, Fahrenkrog B & Aebi U (1999) The nuclear pore complex: from molecular architecture to functional dynamics. *Curr. Opin. Cell Biol.* 11, 391–401.

Weis K (1998) Importins and exportins: how to get in and out of the nucleus. *Trends Biochem. Sci.* 23, 185–189.

Wente SR (2000) Gatekeepers of the nucleus. *Science* 288, 1374–1377.

The Transport of Proteins Into Mitochondria and Chloroplasts

Chen X & Schnell DJ (1999) Protein import into chloroplasts. *Trends Cell Biol.* 9, 222–227.

Cline K & Henry R (1996) Import and routing of nucleus-encoded chloroplast proteins. *Annu. Rev. Cell Dev. Biol.* 12, 1–26.

Dalbey RE & Robinson C (1999) Protein translocation into and across the bacterial plasma membrane and the plant thylakoid membrane. *Trends Biochem. Sci.* 24, 17–22.

Haucke V & Schatz G (1997) Import of proteins into mitochondria and chloroplasts. *Trends Cell Biol.* 7, 103–106.

Jensen RE & Johnson AE (1999) Protein translocation: is Hsp70 pulling my chain? *Curr. Biol.* 9, R779–R782.

Neupert W (1997) Protein import into mitochondria. *Annu. Rev. Biochem.* 66, 863–917.

Rassow J & Pfanner N (2000) The protein import machinery of the mitochondrial membranes. *Traffic* 1, 457–464.

Schwartz MP & Matouschek A (1999) The dimensions of the protein import channels in the outer and inner mitochondrial membranes. *Proc. Natl. Acad. Sci. USA* 96, 13086–13090.

Peroxisomes

Purdue PE & Lazarow PB (2001) Peroxisome biogenesis. *Annu. Rev. Cell Dev. Biol.* 17, 701–752.

Tabak HF, Braakman I & Distel B (1999) Peroxisomes: simple in function but complex in maintenance. *Trends Cell Biol.* 9, 447–453.

The Endoplasmic Reticulum

Adelman MR, Sabatini DD & Blobel G (1973) Ribosome-membrane interaction: nondestructive dissembly of rat liver rough microsomes into ribosomal and membranous components. *J. Cell Biol.* 56, 206–229.

Bergeron JJ, Brenner MB, Thomas DY & Williams DB (1994) Calnexin: a membrane-bound chaperone of the endoplasmic reticulum. *Trends Biochem. Sci.* 19, 124–128.

Bishop WR & Bell RM (1988) Assembly of phospholipids into cellular membranes: biosynthesis, transmembrane movement and intracellular translocation. *Annu. Rev. Cell Biol.* 4, 579–610.

Blobel G & Dobberstein B (1975) Transfer of proteins across membranes. *J. Cell Biol.* 67, 835–851.

Borgese N, Mok W, Kreibich G & Sabatini DD (1974) Ribosomal-membrane interaction: in vitro binding of ribosomes to microsomal membranes. *J. Mol. Biol.* 88, 559–580.

Daleke DL & Lyles JV (2000) Identification and purification of aminophospholipid flippases. *Biochim. Biophys. Acta* 1486, 108–127.

deDuve C (1971) Tissue fractionation—past and present. *J. Cell Biol.* 50 20d–55d.

Deshaies RJ, Sanders SL, Feldheim DA & Schekman R (1991) Assembly of yeast Sec proteins involved in translocation into the endoplasmic reticulum into a membrane-bound multisubunit complex. *Nature* 349, 806–808.

Ellgaard L & Helenius A (2001) ER quality control: towards an understanding at the molecular level. *Curr. Opin. Cell Biol.* 13, 431–437.

Ferguson MA (1992) Colworth Medal Lecture. Glycosylphosphatidylinositol membrane anchors: the tale of a tail. *Biochem. Soc. Trans.* 20, 243–256.

Gething MJ (1999) Role and regulation of the ER chaperone BiP. *Semin. Cell Dev. Biol.* 10, 465–472.

Görlich D, Prehn S, Hartmann E et al. (1992) A mammalian homolog of SEC61p and SECYp is associated with ribosomes and nascent polypeptides during translocation. *Cell* 71, 489–503.

Johnson AE & van Waes MA (1999) The translocon: a dynamic gateway at the ER membrane. *Annu. Rev. Cell Dev. Biol.* 15, 799–842.

Keenan RJ, Freymann DM, Stroud RM & Walter P (2001) The signal recognition particle. *Annu. Rev. Biochem.* 10, 755–775.

Milstein C, Brownlee G, Harrision T & Mathews MB (1972) A possible precursor of immunoglobulin light chains. *Nature New Biol.* 239, 117–120.

Muniz M & Riezman H (2000) Intracellular transport of GPI-anchored proteins. *EMBO J.* 19, 10–15.

Parodi AJ (2000) Role of N-oligosacchanide endoplasmic reticulum processing reactions in glycoprotein folding and degradation. *Biochem. J.* 348, 1–13.

Plemper RK & Wolf DH (1999) Retrograde protein translocation: ERADication of secretory proteins in health and disease. *Trends Biochem. Sci.* 24, 266–270.

Rogers DP & Bankaitis VA (2000) Phospholipid transfer proteins and physiological functions. *Int. Rev. Cytol.* 197, 35–81.

Römisch K (1999) Surfing the Sec61 channel: bidirectional protein translocation across the ER membrane. *J. Cell Sci.* 112, 4185–4191.

Sidrauski C, Chapman R & Walter P (1998) The unfolded protein response: an intracellular signalling pathway with many surprising features. *Trends Cell Biol.* 8, 245–249.

Simon SM & Blobel G (1991) A protein-conducting channel in the endoplasmic reticulum. *Cell* 65, 371–380.

Staehelin LA (1997) The plant ER: a dynamic organelle composed of a large number of discrete functional domains. *The Plant Journal* 11, 1151–1165.

von Heijne G (1995) Membrane protein assembly: rules of the game. *Bioessays* 17, 25–30.

Yan Q & Lennarz WJ (1999) Oligosaccharyltransferase: a complex multisubunit enzyme of the endoplasmic reticulum. *Biochem. Biophys. Res. Commun.* 266, 684–689.